



DISEÑO ÓPTIMO DE EXPERIMENTOS  
PARA MODELOS DE  
ECUACIONES SIMULTÁNEAS

Tesis Doctoral  
Programa de doctorado FISYMAT  
Departamento de Matemáticas

Víctor M. Casero Alonso  
*Director:* Prof. D. Jesús F. López Fidalgo

Ciudad Real, 2014



A Loli,  
Irene, María y Víctor.

Esta tesis no estaría viendo la luz si el Profesor Jesús López Fidalgo no hubiese puesto su confianza en mí. Le agradezco además sus numerosas horas de inestimable dedicación aportándome sus grandes conocimientos de la materia, su gran paciencia y su sabia dirección.

Y tampoco estaría escribiendo estas líneas sin la confianza, ayuda, estímulo y amor de mi esposa (y coautora moral) Loli que me ha apoyado, animado y asistido en el trabajo y me ha quitado ocupaciones y preocupaciones con la familia que hemos ido ampliando: Irene, María y Víctor. Valgan estas pocas palabras para agradecerles todos sus esfuerzos, atenciones, ánimos... En mi corazón hay muchas más que no consigo expresar aquí.

También quiero agradecer a mis padres Ireneo y Antonia el que me dieran la oportunidad de embarcarme en estos estudios 'que no acaban nunca'. Y a mi hermana M<sup>a</sup> Isabel por su inagotable cariño. A mis suegros Fermín y Juliana que también han aportado su granito de arena a este trabajo y que Fermín ve ya con otros ojos desde el cielo.

Tampoco quiero olvidarme de mis compañeros del grupo de diseño óptimo de experimentos de la UCLM (Raúl, Licesio, Mariano, Mercedes Fernández, Mercedes Rodríguez, Santiago, Elvira e Irene) que, aunque un poco disgregados, hemos hecho piña en numerosos congresos y eventos y me han transmitido sus conocimientos, su ánimo y su apoyo. Amplío el agradecimiento a los compañeros de diseño óptimo de Salamanca, Tenerife, Pamplona, Zaragoza, Almería... del nodo Castilla-La Mancha de Biostatnet, y a tantos colegas de trabajo conocidos en congresos, reuniones, escuelas de verano..., por sus constructivas preguntas, sabios comentarios, oportunas sugerencias y tantos buenos momentos que hemos compartido. Especial agradecimiento merecen por mi parte los profesores Werner Müller y Milan Stehlík por su acogida y su trabajo durante mi estancia en la Johannes Kepler University of Linz (Austria).

Que me perdonen aquellos a los que no nombro, y a los que parece que no agradezco, pero tengo en la cabeza. Uno es limitado, como los listados, y para ello mejor dejarlo así.



# Índice general

<b>Índice general</b>	<b>I</b>
<b>Índice de Figuras</b>	<b>III</b>
<b>Índice de Tablas</b>	<b>V</b>
<b>Introducción</b>	<b>VII</b>
<b>1. Teoría del Diseño Óptimo de Experimentos</b>	<b>1</b>
1.1. Diseño clásico de experimentos . . . . .	2
1.2. Modelos lineales . . . . .	3
1.3. Etapas del diseño óptimo de experimentos . . . . .	4
1.4. Diseños exactos y diseños aproximados . . . . .	6
1.5. Matriz de información . . . . .	6
1.6. Corolario del Teorema de Caratheodory . . . . .	7
1.7. Funciones criterio . . . . .	9
1.8. Teorema General de Equivalencia . . . . .	12
1.9. Eficiencia de un diseño . . . . .	15
1.10. Modelos no lineales . . . . .	16
1.11. Críticas al diseño óptimo de experimentos . . . . .	17
1.12. Breve historia del diseño óptimo . . . . .	18
<b>2. Modelos de ecuaciones simultáneas</b>	<b>23</b>
2.1. Modelo estructural de ecuaciones simultáneas . . . . .	23
2.2. Modelo reducido de ecuaciones simultáneas . . . . .	25
2.3. Estimación de los parámetros . . . . .	26
<b>3. Diseños óptimos en modelos de ecuaciones simultáneas</b>	<b>27</b>
3.1. Diseños en modelos triangulares de ecuaciones simultáneas . .	27
3.2. Criterio robusto de composición integral . . . . .	36
3.3. Diseños óptimos con restricciones de costes . . . . .	40
<b>4. Conclusiones y líneas de trabajo futuro</b>	<b>51</b>

<b>Bibliografía</b>	<b>59</b>
<b>Anexos</b>	<b>71</b>
A.1 Experimental designs in triangular simultaneous equations models	
A.2 Robust integral compounding criteria for trend and correlation structures	
A.3 Optimal designs subject to cost constraints in simultaneous equations models	

# Índice de figuras

3.1. Valores de $p^*$ para los diseños CR D-óptimos (izquierda) y SE D-óptimos (derecha) para $\chi = \{0, 1\}$ y $(r_0, r_1) \in [0, 1] \times [0, 1]$ .	31
3.2. Diseños D-óptimos para $\chi = [0, 1]$ para los modelos CR (izquierda) y SE (derecha).	32
3.3. Diseños SE D-óptimos para $\chi = [0, T]$ cuando $T = 1$ (izquierda) y $T = 100$ (derecha).	33



# Índice de tablas

3.1. Diseños exactos óptimos para estimar los parámetros estructurales del ejemplo numérico. Los Casos 1, 2 y 3 son soluciones directas del problema (3.4). P indica la solución obtenida con el procedimiento establecido en esta sección. . . . .	46
3.2. Costes unitarios redondeados. . . . .	46
3.3. Diseños L-óptimos para estimar los parámetros estructurales.	47



# Introducción

Esta tesis doctoral se enmarca en la teoría de Diseño Óptimo de Experimentos, que tiene por objeto la búsqueda del diseño que permite obtener la mayor información posible de un fenómeno con la mínima experimentación. El objetivo concreto perseguido en este trabajo es encontrar diseños óptimos cuando se consideran modelos de ecuaciones simultáneas. El uso de estos modelos se extiende desde la Economía hasta la Ingeniería pasando por las Finanzas, la Medicina y el Marketing entre otros campos.

## Estado del arte

Las referencias bibliográficas sobre diseño óptimo de experimentos para modelos de ecuaciones simultáneas son escasas. Sin embargo, el diseño óptimo de experimentos por un lado, y los modelos de ecuaciones simultáneas por otro han sido ampliamente estudiados. Parafraseando a Draper y Pukelsheim (1996) la bibliografía de cada uno de los temas podría ocupar una enciclopedia entera. Un número especial del *Journal of Econometrics* de 1979 reconoció la importancia del diseño de experimentos en Econometría. En él, Aigner (1979) estudia la asignación óptima de observaciones para regresión multivariante y Conlisk (1979) considera un experimento con los precios de la gasolina utilizando un modelo de ecuaciones simultáneas. Este número especial contribuyó al aumento de experimentos en Economía con algunos artículos significativos como los de Aigner y Balestra (1988) que realizan experimentos con los precios de la electricidad o Madden (1993 y 1995) que realiza experimentos en la red australiana de telefonía. Más recientemente encontramos los experimentos de elección combinada enfocados desde el punto de vista del diseño óptimo por Kessels et al. (2009) o experimentos sociales como el de Hahn, Keisuke y Dean (2011).

Del estudio de la literatura se desprende que el primer artículo que trató conjuntamente el diseño óptimo y las ecuaciones simultáneas fue el de Conlisk (1979). En él se propone ampliar la teoría de diseños óptimos para modelos de regresión a modelos de ecuaciones simultáneas de la forma natural, esto es, minimizando una función escalar de la matriz de covarianzas de los estimadores de los parámetros estructurales del modelo. Posteriormente, Papakyriazis (1986) sugirió la combinación de estimación secuencial y es-

trategias de control del diseño, para tratar el problema de la dependencia de dicha matriz de covarianzas respecto de los parámetros.

### **Novedad de esta tesis**

Esta tesis aporta varias novedades, bien desde el prisma de las ecuaciones simultáneas, o bien desde el diseño óptimo. Por lo mencionado anteriormente, el tema de esta tesis ha sido un reto en cuanto que entra en un terreno muy poco explorado. Por un lado la economía no es un área propensa a considerar la experimentación en la base de su investigación. Por otro lado la falta de resultados relativos a diseño de experimentos con varios modelos simultáneos y correlacionados no facilita el avance en este tema. Además la dificultad técnica no es menor si se tiene en cuenta que en diseño óptimo de experimentos en cuanto se complican las hipótesis deja de haber resultados explícitos.

Para empezar, hemos considerado el problema de diseño óptimo con variables no controlables considerado por López-Fidalgo y Garcet-Rodríguez (2004) que desarrollan la teoría de diseños condicionalmente restringidos, en adelante CR<sup>1</sup>. La novedad aquí es utilizar un modelo de ecuaciones simultáneas, en adelante SE, para afrontar el problema. Se permite así no fijar dicha variable sino considerarla aleatoria formando un sistema de ecuaciones con el modelo original. La variable no controlable, explicativa en la primera ecuación, también será aleatoria existiendo correlación contemporánea entre los términos de error de las ecuaciones. Esto permite un enfoque más general al poder recoger interacciones entre variables que no pueden aparecer en un modelo uniecuacional. Las peculiaridades del modelo considerado en el caso práctico nos llevaron a utilizar un modelo triangular de ecuaciones simultáneas.

En el desarrollo anterior la matriz de información del modelo SE contiene a la matriz de información del modelo CR. De aquí surge el segundo problema investigado. Dado que considerar ambos planteamientos al mismo nivel no parece apropiado, se construye un criterio de composición integral de matrices de información con el que obtener diseños robustos, con un adecuado compromiso entre ambas partes del modelo. El método, basado en optimizar con respecto a una medida el criterio de composición integral, se ilustra con el ejemplo del modelo SE anterior.

Por último, el problema de diseño en un modelo de ecuaciones simultáneas con restricciones de costes que plantea Conlisk (1979) conduce a un problema de programación no lineal continua. Aquí la novedad viene en forma de actualización del problema desde el punto de vista del diseño óptimo. Siguiendo a Cook y Wong (1994) afrontamos el problema considerando una función criterio convexa que combina un criterio estadístico con una función de costes. El resultado es la definición de un procedimiento iterativo

---

<sup>1</sup>todas las siglas utilizadas en esta memoria provienen de los términos en inglés

eficiente, basado en el algoritmo multiplicativo, para obtener diseños óptimos que resuelvan el problema con restricciones. Además el procedimiento permite rebajar el tiempo de cómputo manteniendo las eficiencias de los diseños, incluso mejorándolas tomando como referencia un ejemplo ilustrativo de Conlisk (1979).

### **Estructura de la memoria**

El trabajo desarrollado se estructura en cuatro capítulos. En el primero se describe de forma muy somera el diseño clásico de experimentos y los modelos lineales para introducirse en la teoría general de diseño óptimo de experimentos. Se describen las principales características y herramientas de esta teoría, entre otras, la distinción entre diseños exactos y aproximados, la matriz de información asociada a un diseño, los criterios de optimización y el Teorema General de Equivalencia. En las últimas secciones del capítulo se introducen brevemente los modelos no lineales, se comentan las principales críticas que ha recibido la teoría de diseño óptimo de experimentos y se señalan algunos de los hitos históricos más importantes.

El Capítulo 2 sirve para introducir el modelo de ecuaciones simultáneas, tanto su especificación estructural como la forma reducida del modelo. También se describen los principales métodos de estimación de los parámetros, centrándose en los denominados de información completa que buscan la estimación conjunta de todos los parámetros frente a los de información limitada que, de forma errónea, no aprovechan para la estimación la estructura de correlaciones del sistema.

En el Capítulo 3 se encuentran los resultados de los tres problemas planteados considerando modelos de ecuaciones simultáneas. El Capítulo 4 contiene las principales conclusiones de esta tesis así como las futuras líneas de investigación.

Como fruto ya constatable de esta tesis se encuentran dos artículos publicados en revistas incluidas en el Journal Citation Reports que se incluyen como Anexo. También se incluye en formato artículo los resultados del tercer problema, que se han enviado para su publicación.

Previo al Anexo se ha incluido la bibliografía utilizada durante todo el proceso de elaboración de esta tesis. En ella se incluyen tanto las publicaciones de diseño óptimo y/o modelos de ecuaciones simultáneas, teóricos y/o aplicados citadas en el texto, como las referencias de los artículos de la tesis.

Los cálculos necesarios, tanto simbólicos como numéricos se han realizado con el programa *Mathematica*<sup>®</sup>. El resultado de esos cálculos son los gráficos y datos que se intercalan en el texto. También se ha usado el software libre BAZI de la universidad de Augsburgo que permite aplicar distintos métodos de redondeo.



## Capítulo 1

# Teoría del Diseño Óptimo de Experimentos

Es habitual la necesidad de realizar experimentos para poder inferir de los datos experimentales información valiosa para el conocimiento de una realidad. En este punto surge la cuestión del número de experimentos a realizar y las condiciones experimentales en que deben hacerse tales experimentos. Normalmente la respuesta a estas cuestiones no es única y depende de los objetivos perseguidos. Surge así, de manera natural, la búsqueda del diseño óptimo de experimentos, esto es, el diseño que permite obtener la mayor información posible con la mínima experimentación.

Antes de introducir las principales características de la teoría general del diseño óptimo se describe de forma muy somera el diseño clásico de experimentos, centrando la atención en el procedimiento estadístico, especialmente en la elección del modelo y destacando su importancia al ahorrar tiempo, dinero y riesgos. En el diseño óptimo de experimentos interesa el modelo de regresión, en el que se conoce de antemano el conjunto de puntos observables. No obstante, la elección del modelo para la búsqueda del diseño óptimo es un problema abierto que no tiene una solución general, como veremos en el primer problema al plantear dos modelizaciones para afrontarlo. Interesa buscar un diseño que dé estimadores precisos para el modelo elegido, y que simultáneamente proporcione protección frente a modelos inadecuados. Un buen diseño en un aspecto puede no serlo tanto en otro. Por todo ello, como punto de partida del desarrollo de la teoría de diseño óptimo, se empieza describiendo el modelo lineal.

El objetivo último del diseño óptimo es seleccionar las condiciones experimentales siendo importante la distinción entre diseños exactos y aproximados. En los primeros se conoce de antemano el número  $n$  de observaciones a realizar consistiendo el diseño óptimo en elegir los puntos del espacio de diseño donde realizar las observaciones. En los diseños aproximados, teóricamente, puede asignarse un número decimal de observaciones a los puntos del

espacio de diseño. En la práctica es necesario un redondeo del diseño aproximado. El uso de diseños aproximados es muy conveniente para demostrar ciertas propiedades matemáticas.

Para determinar el número de observaciones en cada punto es necesaria la matriz de información asociada a un diseño. Ésta es la herramienta básica del diseño óptimo de experimentos pues en ella se basan los criterios de optimización. También se describen otras herramientas útiles para la obtención de diseños óptimos. El corolario del Teorema de Caratheodory permite restringirse a diseños con soporte finito. El Teorema General de Equivalencia, GET, que demuestra la equivalencia entre los criterios D- y G-óptimo en primera instancia, proporciona condiciones para comprobar la optimalidad de un diseño basándose en la derivada direccional de una función criterio. Además el GET es la base para la mayoría de algoritmos de cálculo de diseños óptimos, así como para calcular la eficiencia de un diseño particular. En el caso en el que la función criterio sea diferenciable se define la función de sensibilidad que simplifica el enunciado del GET.

En los últimos apartados del capítulo se abordan brevemente los modelos no lineales, se comentan las principales críticas que ha recibido la teoría de diseño óptimo de experimentos y se señalan algunos de los hitos históricos más importantes.

Las referencias principales para la elaboración de este capítulo han sido los libros de Pázman (1986) y Atkinson y Donev (1992) junto con el manual de López-Fidalgo (2011). A ellos puede acudir para encontrar más detalles y demostraciones.

## 1.1. Diseño clásico de experimentos

### Procedimiento estadístico

Todo procedimiento estadístico de diseño de experimentos debe empezar por la elección del modelo, en función de los objetivos del experimento. Cada modelo lleva aparejado un diseño experimental, entendiéndose éste como la elección de las condiciones experimentales en las que realizar los experimentos así como una primera idea del número de experimentos a realizar en función de diversas restricciones, generalmente presupuestarias.

En ocasiones se elige directamente un procedimiento estadístico sin pensar en el modelo que le precede. Sin embargo, al elegir el procedimiento se está eligiendo implícitamente el modelo. Así, si se realiza un análisis de la varianza, ANOVA, se está considerando implícitamente un modelo donde la variable respuesta es cuantitativa y las variables explicativas son cualitativas. O si se realiza un análisis de regresión se está considerando el modelo en el que tanto la variable respuesta como las explicativas son cuantitativas.

Obtenidos los datos experimentales llega la etapa de su análisis. Previamente quizá sea necesaria alguna transformación de los mismos para un

análisis más adecuado. Analizada la información experimental se utiliza la teoría estadística para su interpretación y la toma de decisiones.

### La importancia de diseñar un experimento

Un diseño experimental es fruto del pensamiento previo a la acción. Bien planteado, el diseño de experimentos permite ahorrar tiempo, dinero y riesgos. Y por supuesto, un buen diseño de experimento permite realizar un análisis correcto, lo que no significa que el análisis de un experimento siempre sea correcto.

El diseño de experimentos clásico se basa en tres principios: la *aleatoriedad*, la *replicación* y los *factores bloque*. La *aleatoriedad* consiste en la realización o medición de los experimentos de modo aleatorio para evitar efectos no deseados, como por ejemplo el efecto aprendizaje. La *replicación* consiste en construir o utilizar unidades experimentales distintas con las mismas condiciones experimentales nominales. Suele confundirse con medidas repetidas. El *diseño en bloques* se utiliza en aquellos casos en los que una variable no es de interés en el estudio pero puede afectar a la respuesta, por lo que se toma en consideración en el análisis y así se eliminan “ruidos” en la explicación de la variabilidad.

### Directrices para el diseño de experimentos

Siguiendo a Montgomery (2001) podemos señalar las siguientes etapas en el diseño de experimentos: planteamiento preexperimental, elección del diseño experimental, realización del experimento, análisis estadístico y conclusiones y recomendaciones.

El planteamiento preexperimental incluiría el planteamiento del problema, la elección del modelo, es decir, la selección de los factores (controlables, no controlables y ruido), niveles y rangos y la selección de la(s) variable(s) respuesta. Para la elección del diseño experimental cabe citar, entre otros muchos, diseños factoriales, diseños factoriales fraccionales, diseños anidados o jerárquicos, diseños split-plot, diseños secuenciales y adaptativos, diseños con nombre propio... En la realización del experimento es importante la monitorización del proceso. Tras la obtención de los datos vendría el análisis estadístico de los mismos. Por último, se obtendrían las conclusiones y recomendaciones.

## 1.2. Modelos lineales

En general intentar explicar una realidad, habitualmente compleja, mediante un modelo suele ser difícil, cuando no utópico. Es lógico buscar modelos que aproximen lo mejor posible, con un margen de error aceptable, tal realidad. En ocasiones tales modelos son no lineales, tanto en las variables

como en los parámetros, y en otras, un modelo lineal puede explicar suficientemente bien la realidad objeto de estudio. En el desarrollo de la teoría del diseño óptimo de experimentos se han considerado primero modelos lineales y después modelos no lineales. Seguiremos aquí este planteamiento.

Se entiende por **modelo lineal** aquél que es lineal en los parámetros, aunque no lo sea en las variables. Esta definición se aplica tanto a los modelos de regresión uniecuacionales como a los modelos de ecuaciones simultáneas en los que se centra esta tesis.

### 1.3. Etapas del diseño óptimo de experimentos

En el diseño óptimo de experimentos, como caso particular del diseño de experimentos, se dan las mismas etapas descritas anteriormente. No obstante, existen algunas particularidades que conviene citar.

#### Elección del modelo

En el diseño óptimo de experimentos estamos interesados en el modelo de regresión, en el que se conoce de antemano el conjunto de puntos observables. La elección del modelo para la búsqueda del diseño óptimo es un problema abierto que no tiene una solución general. Interesa buscar un diseño que dé estimadores precisos para el modelo elegido, y que simultáneamente proporcione protección frente a modelos inadecuados. Un buen diseño en un determinado aspecto puede no serlo tanto en otro (López-Fidalgo, 2011).

Por tanto, el diseño óptimo de un experimento se desarrolla en torno a la elección del modelo de regresión, que para empezar consideraremos lineal. Una forma útil de escribirlo es

$$E(y|x) = f^T(x)\theta, \quad \text{Var}(y|x) = \sigma^2, \quad x \in \chi \quad (1.1)$$

donde  $y$  es la variable respuesta,  $x$  es la variable o variables que representan las condiciones experimentales y  $f$  es una función continua conocida, con tantas componentes linealmente independientes como  $\theta$ , el vector de parámetros desconocidos. Con  $\text{Var}(y|x) = \sigma^2$  se expresa que las observaciones de la variable respuesta,  $(y_1, \dots, y_n)^T$ , se suponen no correlacionadas y con varianza constante  $\sigma^2$ . El caso de observaciones correlacionadas supone otro enfoque en el que no se entrará aquí (véase p. ej. Müller y Stehlík, 2009).

#### Diseño óptimo

El segundo paso es seleccionar las condiciones experimentales  $x$  sobre el **espacio de diseño**  $\chi$ . Es decir, obtener el diseño óptimo. En la práctica el espacio  $\chi$  va a ser un subconjunto compacto sobre un espacio Euclídeo.

En la obtención del diseño óptimo influyen la elección del criterio de optimización, que trataremos en la sección 1.7, confrontando la teoría con

la situación real, y la utilización de un algoritmo para el cálculo del diseño óptimo. Si el algoritmo es numérico debe poder medirse la distancia que separa el diseño óptimo de los diseños que se van obteniendo en cada paso. La obtención de diseños óptimos no es una tarea fácil.

Algunos ejemplos de modelos lineales y espacios de diseño son:

- $y = \alpha_0 + \alpha_1 x + \varepsilon$ 
  - $f(x) = (1, x)^T$ ,  $x \in \chi = [a, b]$
- $y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \varepsilon$ 
  - $f(x) = x = (x_1, x_2, x_3)^T$ ,  $x \in \chi = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$
- $y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \varepsilon$ 
  - $f(x) = (1, x, x^2)^T$ ,  $x \in \chi = [a, b]$
- ANOVA de un factor y 3 niveles,  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ 
  - $f(x_1, x_2) = (1, x_1, x_2)^T$ ,  $x = (x_1, x_2) \in \{(0, 1), (1, 0), (1, 1)\}$

### Información experimental, estimación e inferencia

El siguiente paso sería obtener las respuestas  $y_1, \dots, y_n$  para las condiciones experimentales  $x_1, \dots, x_n$  seleccionadas.

A continuación se estimarían los parámetros desconocidos del modelo  $\theta_1, \dots, \theta_m, \sigma^2$  mediante, por ejemplo, el método de mínimos cuadrados, o el de máxima verosimilitud y se harían las inferencias correspondientes.

Para expresar los estimadores de los parámetros es conveniente escribir el modelo en forma matricial

$$\begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} f_1(x_1) & \dots & f_m(x_1) \\ \dots & \dots & \dots \\ f_1(x_n) & \dots & f_m(x_n) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \dots \\ \theta_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix} \Rightarrow Y = X\Theta + \varepsilon,$$

donde  $\varepsilon$  es el término de error experimental.

Con esta notación, el estimador por el método de mínimos cuadrados, LSE, es

- Para el caso ordinario, matriz de covarianzas  $\Sigma_Y = I$

$$\hat{\Theta}_{LSE} = (X^T X)^{-1} X^T Y,$$

siendo su matriz de covarianzas

$$\Sigma_{\hat{\Theta}_{LSE}} = \sigma^2 (X^T X)^{-1}$$

- Para el caso generalizado,  $\Sigma_Y$  general

$$\hat{\Theta}_{LSE} = (X^T \Sigma_Y^{-1} X)^{-1} X^T \Sigma_Y^{-1} Y,$$

siendo su matriz de covarianzas

$$\Sigma_{\hat{\Theta}_{LSE}} = \sigma^2 (X^T \Sigma_Y^{-1} X)^{-1}$$

El Teorema de Gauss–Markov establece una muy buena propiedad de los estimadores mínimo cuadráticos.

**Teorema 1.1 (Teorema de Gauss–Markov)**

*Si  $x$  es una variable determinista, i.e. no aleatoria, independiente de  $\varepsilon$  entonces el estimador de mínimos cuadrados,  $\hat{\Theta}_{LSE}$ , es el estimador lineal insesgado de mínima varianza.*

*Por otro lado, bajo el supuesto de normalidad de  $Y$ ,  $\hat{\Theta}_{LSE} = \hat{\Theta}_{MLE}$ , donde  $MLE$  denota el estimador de máxima verosimilitud.*

## 1.4. Diseños exactos y diseños aproximados

El objetivo por tanto del diseño de experimentos es elegir las condiciones experimentales  $x_1, \dots, x_n$ , es decir, elegir  $n$  puntos del espacio de diseño  $\chi$ . Se dice entonces que se tiene un **diseño exacto de tamaño  $n$** , donde  $n$  es conocido de antemano.

Algunas de estas condiciones experimentales podrían repetirse, por lo que es natural considerar una medida de probabilidad discreta  $\xi_n(x) = n_x/n$  siendo  $n_x$  el número de observaciones realizadas en  $x$ . Generalizando la medida de probabilidad discreta de los diseños exactos a cualquier medida de probabilidad  $\zeta$  sobre el espacio de Borel generado por los conjuntos abiertos de  $\chi$  se llega a los **diseños aproximados o asintóticos** (Kiefer, 1959). El conjunto de diseños aproximados se denota por  $\Xi$  y es convexo.

La generalización de los diseños exactos a los diseños aproximados viene a relajar la condición de realizar un número entero de observaciones en el punto  $x$ . En la práctica no sería posible realizar un número decimal de observaciones  $n_x = n\xi(x)$  (véase el ejemplo de la sección 1.6). No obstante, estos diseños son muy convenientes para demostrar ciertas propiedades. Si  $n$  es suficientemente grande podría aproximarse un diseño aproximado a uno exacto. Aunque, como veremos en la Sección 3.1 con una pseudo–paradoja, puede que siempre difieran. Imhof, López–Fidalgo y Wong (2001) recomiendan no utilizar diseños aproximados para  $n$  pequeño.

## 1.5. Matriz de información

A partir de la matriz de covarianzas de los estimadores de mínimos cuadrados definida anteriormente y considerando la medida de probabilidad discreta  $\xi_n(x)$  podemos escribir

$$\Sigma_{\hat{\Theta}} = \sigma^2(X^T X)^{-1} = \sigma^2 n^{-1} M^{-1}(\xi_n)$$

donde  $M(\xi_n)$  es la **matriz de información** que se define para un diseño exacto como

$$M(\xi_n) = \sum_{x \in \chi} f(x) f^T(x) \xi_n(x)$$

donde  $f$  queda definida en (1.1). Esta matriz coincide, salvo constantes, con la matriz de información de Fisher.

En el caso de diseños aproximados la matriz de información se define como

$$M(\zeta) = \int_{\mathcal{X}} f(x)f^T(x)\zeta(dx).$$

Si la medida de probabilidad  $\zeta$  es discreta y  $\xi$  es su función de probabilidad asociada

$$M(\xi) = \sum_{x \in \mathcal{X}} f(x)f^T(x)\xi(x)$$

que coincide con la definición dada anteriormente para el caso particular de diseños exactos.

Si  $\zeta$  es absolutamente continua, existe su función de densidad asociada,  $\xi$ , y entonces

$$M(\xi) = \int_{\mathcal{X}} f(x)f^T(x)\xi(x)dx.$$

Definidas así, el conjunto de matrices de información  $\mathcal{M}$ , es convexo y compacto.

### Propiedades de las matrices de información

1. Son simétricas.
2. Son semidefinidas positivas.
3. Son singulares si el número de puntos distintos del diseño es inferior al número de parámetros,  $m$ .

### Demostración

1. Por construcción.
2. Puede verse, por ejemplo, por ser inversas de las matrices de covarianzas.
3. Si el número de puntos distintos del diseño es inferior al número de parámetros, habrá al menos dos columnas iguales (o filas por ser simétricas).

■

## 1.6. Corolario del Teorema de Caratheodory

El Teorema de Caratheodory permite demostrar un resultado fundamental en la teoría de diseños óptimos.

**Teorema 1.2 (Corolario del Teorema de Caratheodory)**

Para cualquier matriz de información siempre existe un diseño asociado con a lo sumo  $m(m+1)/2 + 1$  puntos distintos en su soporte.

**Demostración**

Véase Karlin y Studden (1966). ■

La importancia de este resultado radica en que, en la práctica, podemos restringirnos a diseños óptimos aproximados con soporte finito del tipo

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_k \\ p_1 & p_2 & \dots & p_k \end{array} \right\},$$

donde  $x_i$  son los puntos del soporte y  $p_i = \xi(x_i)$  los pesos del diseño, siendo  $0 \leq p_i \leq 1$  para  $i = 1, \dots, k$  y  $\sum_{i=1}^k p_i = 1$ . Se evita así considerar el conjunto convexo de diseños aproximados  $\Xi$ , mucho más difícil de manejar. Por consiguiente, de ahora en adelante utilizaremos  $\xi$  ó  $\zeta$  como más convenga.

En la práctica se realizarán  $n_i \approx n\xi(x_i)$  experimentos en  $x_i$ , ciñéndose a que  $\sum_i n_i = n$ . Esta elección no es tan simple como redondear dichos valores de alguna manera, aquí existe la restricción de que todas las proporciones han de sumar 1. Consideremos como ejemplo el siguiente diseño

$$\xi = \left\{ \begin{array}{ccc} 0 & 0,5 & 1 \\ 0,2 & 0,4 & 0,4 \end{array} \right\}.$$

Si el número de experimentos a realizar es  $n = 12$  habrá que realizar el 20% de ellos en 0, esto es, 2,4 experimentos y el 40%, 4,8 experimentos en 0,5 y en 1. Las posibles combinaciones, teniendo en cuenta que sólo pueden hacerse 12 experimentos, serían:

- 2 experimentos en 0, 5 experimentos en 0,5 y otros 5 en 1.
- 3 experimentos en 0, 4 experimentos en 0,5 y 5 en 1.
- 3 experimentos en 0, 5 experimentos en 0,5 y 4 en 1.

Dependiendo del modelo y del criterio utilizado unos diseños serán mejores que otros. En Imhof, López-Fidalgo y Wong (2001) se puede encontrar una lista de métodos de redondeo. Durante el desarrollo de esta tesis (véase 3.3) se ha usado el software BAZI<sup>1</sup> creado por Friedrich Pukelsheim y el autodenominado Bazi Team, de la Universidad de Augsburgo, que permite aplicar nueve métodos distintos de redondeo o prorrateo para la representación proporcional en su terminología.

<sup>1</sup>Berechnung von Anzahlen mit Zuteilungsmethoden im Internet/ Calculation of Allocations by Apportionment Methods in the Internet. <https://www.uni-augsburg.de/bazi>

## 1.7. Funciones criterio

En el diseño óptimo de experimentos existen diversos criterios de búsqueda de dicho óptimo. Por ejemplo, el criterio de D-optimización consiste en minimizar el determinante de la inversa de la matriz de información asociada al diseño, o el criterio de A-optimización consiste en minimizar la traza de la inversa de la matriz de información.

Cada criterio,  $\Phi$ , lleva asociada una función que podemos definir como

$$\Phi : \Xi \rightarrow [0, +\infty) \quad \text{ó} \quad \Phi : \mathcal{M} \rightarrow [0, +\infty).$$

La primera definición es más general, la segunda es más manejable y es la que utilizaremos en esta memoria.

Se dice que el diseño  $\xi^*$  es  $\Phi$ -óptimo si minimiza  $\Phi$ ,

$$\xi^* = \arg \min_{\xi \in \Xi} \Phi[M(\xi)]$$

La segunda definición de función criterio conlleva una serie de particularidades. La primera es que debe considerarse como función de  $M^{-1}$ . Además, para una mejor estimación de los parámetros ha de considerarse no decreciente en el sentido del orden de Loewner ( $M \leq N$  si  $N - M$  es semidefinida positiva)

$$M \leq N \Rightarrow \Phi(M) \geq \Phi(N).$$

La función criterio puede definirse de forma global o parcial, dependiendo de si interesan todos o sólo una parte de los parámetros del modelo. Además, interesará también que sea positivamente homogénea, es decir,

$$\Phi(\delta M) = \frac{1}{\delta} \Phi(M), \quad \delta > 0.$$

Por último,  $\mathcal{M}$  está en el cono de matrices definidas no negativas en el espacio Euclídeo de las matrices simétricas. Por consiguiente, tiene sentido la definición habitual de diferenciabilidad.

### Observaciones para criterios globales

1. Un diseño  $\Phi$ -óptimo tiene al menos  $m$  puntos en su soporte, siendo  $m$  el número de parámetros del modelo lineal. Cualquier diseño con menos de  $m$  puntos conduce a una matriz singular.
2. Una propiedad útil para las funciones criterio es que sean convexas,

$$\Phi [(1 - \varepsilon)\zeta + \varepsilon\zeta'] \leq (1 - \varepsilon)\Phi(\zeta) + \varepsilon\Phi(\zeta').$$

3. Para criterios estrictamente convexos siempre existe un diseño  $\Phi$ -óptimo con a lo sumo  $m(m+1)/2$  puntos de soporte. Esto se debe a que el óptimo debe estar en la frontera, cuya dimensión es  $(m(m+1)/2) - 1$ . Aplicando ahí el corolario del Teorema de Caratheodory se tiene el resultado.
4. Si  $\sigma^2(x)$  no es constante, todo lo anterior se puede aplicar a  $\tilde{f}(x) = f(x)/\sigma(x)$ . Así, por simplicidad se considerará en lo sucesivo,  $\sigma(x) = 1$ .
5. Si las observaciones están correlacionadas sólo se pueden obtener diseños exactos.

Se dan a continuación la definición y propiedades principales de los criterios utilizados en la elaboración de esta tesis.

### D-optimización

Es el criterio más popular. Consiste en minimizar el determinante, de aquí la D, de la inversa de la matriz de información. En la práctica, se maximiza el determinante de la matriz de información.

Wald (1943) establece este criterio. Más tarde Kiefer y Wolfowitz (1959) le darán el nombre de D-optimización y extenderán su utilización al modelo de regresión más general.

La función criterio es

$$\Phi_D[M(\xi)] = \begin{cases} \log \det M^{-1}(\xi) = -\log \det M(\xi) & \text{si } \det M(\xi) \neq 0, \\ \infty & \text{si } \det M(\xi) = 0. \end{cases}$$

Y su definición como función positivamente homogénea es

$$\Phi_D[M(\xi)] = \begin{cases} \det M^{-1/m}(\xi) & \text{si } \det M(\xi) \neq 0, \\ \infty & \text{si } \det M(\xi) = 0. \end{cases}$$

### Propiedades

1.  $\Phi_D$  es continua en  $\mathcal{M}$ .
2. La función  $\Phi_D$  es convexa en  $\mathcal{M}$  y estrictamente convexa en  $\mathcal{M}_+$ , el conjunto de matrices de información no singulares.
3. La función  $\Phi_D$  es diferenciable en  $\mathcal{M}_+$ . Su gradiente es

$$\nabla[-\log \det M(\xi)] = -M^{-1}(\xi).$$

4.  $\Phi_D$  proporciona el mínimo volumen del elipsoide de confianza de los parámetros del modelo

$$(\hat{\theta} - \theta)^T (X^T X)^{-1} (\hat{\theta} - \theta) \leq m F_{m, n-m, \gamma} S_R^2 \equiv c^2,$$

dado que el volumen es  $c^m V_m [\det M^{-1}(\xi)]^{1/2}$ , siendo  $V_m$  el volumen de la esfera  $m$ -dimensional de radio unidad.

5.  $\Phi_D$  es invariante ante reparametrizaciones.
6. Siempre se puede conseguir que el número de puntos en el soporte del diseño D-óptimo esté comprendido entre  $m$  y  $m(m+1)/2$ .
7. Es fácil de calcular, lo que es una gran ventaja respecto al resto de criterios.

### G-optimización

Este criterio consiste en minimizar el valor más grande posible de la varianza de la respuesta predicha en  $x$ . La G proviene de *Generalized Variance*. El Teorema General de Equivalencia (véase Sección 1.8) establece la equivalencia entre D- y G-optimización. La función criterio asociada es

$$\Phi_G[M(\xi)] = \begin{cases} \max_{x \in X} f^T(x)M^{-1}(\xi)f(x) & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

#### Propiedades

1.  $\Phi_G$  es continua en  $\mathcal{M}$ .
2.  $\Phi_G$  es convexa en  $\mathcal{M}$  y estrictamente convexa en  $\mathcal{M}_+$ .

### L-optimización

El criterio de L-optimización consiste en minimizar la traza del producto de la inversa de la matriz de información por una matriz  $L$ , que refleja el interés del experimentador. Su función criterio asociada fue definida por Atwood (1976)

$$\Phi_L[M(\xi)] = \begin{cases} \text{tr}LM^{-1}(\xi) & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

donde  $L$  es una matriz definida positiva del mismo orden que  $M$ .

#### Propiedades

1.  $\Phi_L$  es continua en  $\mathcal{M}$ .
2.  $\Phi_L$  es convexa en  $\mathcal{M}$  y estrictamente convexa en  $\mathcal{M}_+$ .
3.  $\Phi_L$  es diferenciable en  $\mathcal{M}_+$ . Su gradiente es

$$\nabla[\text{tr}LM^{-1}(\xi)] = -M^{-1}(\xi)LM^{-1}(\xi).$$

### Relación entre L- y A-optimización

El criterio de A-optimización es un caso particular de L-optimización, tomando  $L$  como la matriz identidad,

$$\Phi_A[M(\xi)] = \begin{cases} \frac{1}{m} \text{tr} M^{-1}(\xi) & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

Equivale a minimizar la media, de aquí la A (average), de las varianzas de los estimadores de los parámetros.

### Otros criterios

Existen más criterios de optimización entre los que podemos citar  $c$ -,  $E$ -,  $\Phi_p$ -,  $L_p$ -,  $I$ -,  $MV$ -optimización...

## 1.8. Teorema General de Equivalencia

El Teorema General de Equivalencia, GET, se puede considerar el teorema más importante del diseño óptimo de experimentos. Kiefer y Wolfowitz (1960) fueron los precursores de este teorema siendo los primeros en demostrar la equivalencia de los criterios D-óptimo y G-óptimo. Esto es, utilizando ambos criterios siempre se llega al mismo diseño óptimo.

Whittle (1973) y Kiefer (1974) generalizan el GET para funciones criterio más generales, basándose en la comprobación de las derivadas direccionales. Aquí se va a proporcionar el enunciado de Whittle pero antes se ha de introducir el concepto de derivada direccional.

### Derivada direccional

Ya se ha comentado que tiene sentido hablar de diferenciabilidad de diseños o matrices de información. En este contexto se define la **derivada direccional** de  $\Phi$  en  $\zeta$  en la dirección  $\zeta'$  como (definición de Frechét)

$$\partial\Phi(\zeta, \zeta') = \lim_{\varepsilon \rightarrow 0^+} \frac{\Phi[(1 - \varepsilon)\zeta + \varepsilon\zeta'] - \Phi(\zeta)}{\varepsilon}.$$

Realmente la dirección de la derivada direccional sería  $\zeta' - \zeta$ .

### Propiedades

1.  $\partial\Phi(\zeta, \zeta) = 0$ .
2. Si  $\Phi$  es convexa,  $\partial\Phi(\zeta, \zeta')$  siempre existe y es finita o  $+\infty$ .
3.  $\partial\Phi(\zeta, \zeta') \leq \Phi(\zeta') - \Phi(\zeta)$ .

### Demostración

1. Obvio por la definición.
2. Dado que  $\frac{\Phi[(1-\varepsilon)\zeta + \varepsilon\zeta'] - \Phi(\zeta)}{\varepsilon}$  crece con  $\varepsilon \in (0, 1)$ . Si  $0 < \varepsilon_1 < \varepsilon_2 < 1$ ,

$$\begin{aligned} \Phi[(1-\varepsilon_1)\zeta + \varepsilon_1\zeta'] - \Phi(\zeta) &= \Phi\left\{\frac{\varepsilon_1}{\varepsilon_2}[(1-\varepsilon_2)\zeta + \varepsilon_2\zeta'] + \left(1 - \frac{\varepsilon_1}{\varepsilon_2}\right)\zeta\right\} - \Phi(\zeta) \\ &\leq \frac{\varepsilon_1}{\varepsilon_2}\Phi[(1-\varepsilon_2)\zeta + \varepsilon_2\zeta'] + \left(1 - \frac{\varepsilon_1}{\varepsilon_2}\right)\Phi(\zeta) - \Phi(\zeta) \\ &= \frac{\varepsilon_1}{\varepsilon_2}\{\Phi[(1-\varepsilon_2)\zeta + \varepsilon_2\zeta'] - \Phi(\zeta)\} \end{aligned}$$

3.

$$\begin{aligned} \partial\Phi(\zeta, \zeta') &= \lim_{\varepsilon \rightarrow 0^+} \frac{\Phi[(1-\varepsilon)\zeta + \varepsilon\zeta'] - \Phi(\zeta)}{\varepsilon} \\ &\leq \lim_{\varepsilon \rightarrow 0^+} \frac{(1-\varepsilon)\Phi(\zeta) + \varepsilon\Phi(\zeta') - \Phi(\zeta)}{\varepsilon} \\ &= \Phi(\zeta') - \Phi(\zeta) \end{aligned}$$

■

Por otro lado, si  $\Phi$  es diferenciable, usando el desarrollo en serie de Taylor

$$\partial\Phi(M, N) = \text{tr}[\nabla\Phi(M)(N - M)], \quad (1.2)$$

donde  $\text{tr}$  indica la traza y  $\nabla$  el gradiente (véase Pázman, 1986). Como corolario  $\partial\Phi(M, N)$  es lineal en el segundo argumento, lo que es una condición necesaria y suficiente para la diferenciabilidad. En un espacio más general, por ejemplo,  $\Xi$ , una función se denomina diferenciable en  $\zeta$  si la derivada direccional es lineal en las direcciones:

$$\partial\Phi(\zeta, \zeta') = \int_{\mathcal{X}} \partial\Phi(\zeta, \zeta_x)\zeta'(dx).$$

Ya se está en condiciones de enunciar la versión de Whittle (1973) del GET.

**Teorema 1.3 (Teorema General de Equivalencia, GET)**

*Si  $\Phi$  es convexa, las tres condiciones siguientes son equivalentes*

- a.  $\Phi(\zeta^*) = \min_{\zeta \in \Xi} \Phi(\zeta)$
- b.  $\inf_{\zeta' \in \Xi} \partial\Phi(\zeta^*, \zeta') = \max_{\zeta \in \Xi} \inf_{\zeta' \in \Xi} \partial\Phi(\zeta, \zeta')$
- c.  $\inf_{\zeta' \in \Xi} \partial\Phi(\zeta^*, \zeta') = 0$ .

*Si  $\Phi$  es diferenciable, para diseños con función de densidad o probabilidad  $\xi$ ,  $\partial\Phi(\xi^*, \xi_x) = 0, x \in S_{\xi^*}$ , donde  $S_{\xi} = \{x : \xi(x) > 0\}$  es el soporte del diseño.*

### Demostración

Al ser  $f$  continua y  $\chi$  compacto existe un diseño  $\Phi$ -óptimo,  $\zeta^*$ . Se satisface *a*.

Basta probar que es el único diseño que satisface las tres condiciones.

Por optimalidad  $\partial\Phi(\zeta^*, \zeta') \geq 0$  para cualquier  $\zeta' \in \Xi$  e  $\inf_{\zeta' \in \Xi} \partial\Phi(\zeta^*, \zeta') \geq 0$ , que se alcanza puesto que  $\partial\Phi(\zeta^*, \zeta^*) = 0$ . Por consiguiente se satisface *c*.

Por otro lado,  $\inf_{\zeta' \in \Xi} \partial\Phi(\zeta, \zeta') \leq \partial\Phi(\zeta, \zeta) = 0$  para cualquier diseño  $\zeta$ , y se alcanza al menos para  $\zeta^*$ . Así se satisface *b*.

Por último, sea  $\zeta$  un diseño no óptimo. Entonces existe otro diseño  $\zeta''$  tal que  $\Phi(\zeta'') < \Phi(\zeta)$ . Por ello,  $\partial\Phi(\zeta, \zeta'') \leq \Phi(\zeta'') - \Phi(\zeta) < 0$  e  $\inf_{\zeta' \in \Xi} \partial\Phi(\zeta, \zeta') < 0$ . En conclusión, sólo un diseño óptimo puede satisfacer *b*. y *c*.

Si  $\Phi$  es diferenciable, entonces  $\int_{\chi} \partial\Phi(\zeta^*, \zeta_x) \zeta^*(dx) = \partial\Phi(\zeta^*, \zeta^*) = 0$ , pero  $\partial\Phi(\zeta^*, \zeta_x) \geq 0$ , así la derivada direccional debe anularse en los puntos del soporte. ■

### Observaciones del GET

1. El GET no se cumple, en general, para diseños exactos. Se puede encontrar un ejemplo en Atkinson, Donev y Tobias (2007), página 125.
2. Como  $\mathcal{M}$  es convexo el teorema es aplicable tanto para diseños finitos como para diseños más generales. Por ello, de ahora en adelante, supondremos diseños finitos y utilizaremos el símbolo  $\xi$  (probabilidad) en lugar de  $\zeta$  (medida).
3. El teorema sigue siendo válido para búsquedas restringidas en subconjuntos convexos.

### Función de sensibilidad

Si la función criterio es diferenciable se simplifica el Teorema 1.3 utilizando la **función de sensibilidad** que definimos como

$$\psi(x, \xi) = \partial\Phi[M(\xi), f(x)f^T(x)]$$

Usando (1.2) podemos reescribirla así

$$\psi(x, \xi) = f^T(x)\nabla\Phi[M(\xi)]f(x) - \text{tr}M(\xi)\nabla\Phi[M(\xi)].$$

### Teorema 1.4 (*GET para criterios diferenciables*)

$\xi^*$  es  $\Phi$ -óptimo si y sólo si  $\psi(x, \xi^*) \geq 0$ ,  $x \in \chi$ .

La igualdad se da para  $x \in S_{\xi^*}$ .

### Demostración

Para empezar, podemos expresar la derivada direccional de una matriz de información como suma de derivadas direccionales si  $\xi'$  es finito

$$\partial\Phi [M(\xi), M(\xi')] = \sum_x \partial\Phi [M(\xi), M(\xi_x)] \xi'(x).$$

Por otro lado, del apartado *c.* del Teorema 1.3, se tiene que

$$\partial\Phi [M(\xi^*), f(x)f^T(x)] = \psi(x, \xi^*) \geq 0$$

para cualquier  $x \in \chi$ . ■

Esta versión del GET proporciona un par de condiciones para comprobar la  $\Phi$ -optimalidad de un diseño dado, a saber:

1.  $\psi(z, \xi^*) = 0$ ,  $z \in S_{\xi^*}$
2.  $\left( \frac{\partial\psi(x, \xi^*)}{\partial x} \right)_{x=z} = 0$ ,  $z \in S_{\xi^*} \cap \text{Int}(\chi)$

El GET es la base para la mayoría de algoritmos de cálculo de diseños óptimos, así como para calcular la eficiencia de un diseño particular.

## 1.9. Eficiencia de un diseño

La eficiencia se puede interpretar como una medida de la bondad de un diseño. Se define la eficiencia de un diseño  $\xi$  respecto a un criterio  $\Phi$  como

$$\text{eff}_\Phi(\xi) = \frac{\Phi[M(\xi^*)]}{\Phi[M(\xi)]}, \quad (1.3)$$

donde  $\xi^*$  es el diseño  $\Phi$ -óptimo.

Si el criterio  $\Phi$  es positivamente homogéneo y para  $n$  observaciones un diseño  $\xi$  tiene una eficiencia del 60 %, entonces

$$0,6 = \frac{\Phi[M(\xi^*)]}{\Phi[M(\xi)]} = \frac{\Phi(\sigma^{-2}n\Sigma_{\hat{\theta}}^*)}{\Phi(\sigma^{-2}n\Sigma_{\hat{\theta}})} = \frac{\sigma^2n^{-1}\Phi(\Sigma_{\hat{\theta}}^*)}{\sigma^2n^{-1}\Phi(\Sigma_{\hat{\theta}})} = \frac{\Phi(\Sigma_{\hat{\theta}}^*)}{\Phi(\Sigma_{\hat{\theta}})}.$$

Así que, siendo  $n^*$  el número de observaciones con el diseño  $\xi^*$  y manteniendo  $n$  observaciones con el diseño  $\xi$ , para que sean igual de eficientes

$$1 = \frac{\Phi(\Sigma_{\hat{\theta}}^*)}{\Phi(\Sigma_{\hat{\theta}})} = \frac{\sigma^2n^{*-1}\Phi[M(\xi^*)]}{\sigma^2n^{-1}\Phi[M(\xi)]} = \frac{\sigma^2n^{*-1}}{\sigma^2n^{-1}}0,6,$$

por lo tanto  $n^* = 0,6n$ . Con el diseño óptimo,  $\xi^*$ , sólo serían necesarias el 60 % de las observaciones del diseño  $\xi$ . Esta interpretación es mucho más fácil de entender para cualquier experimentador que los típicos conceptos estadísticos de *p-valor* o *potencia del contraste*.

De manera análoga se puede definir la eficiencia relativa entre dos diseños, que permitirá compararlos. También se utiliza la eficiencia para realizar el análisis de sensibilidad respecto de los parámetros nominales del modelo, cuando éste es necesario.

### 1.10. Modelos no lineales

Entendemos por modelos estadísticos paramétricos no lineales aquellos que no son lineales en los parámetros (véase la Sección 1.2). Su estructura es similar a los modelos lineales en el siguiente sentido. Se dispone de una variable respuesta,  $y$ , junto con una o varias variables explicativas,  $x$ . El modelo asociado se supone que proviene de una familia de distribuciones con función de densidad  $h(y|x, \theta)$ . Con este planteamiento se está considerado cualquier modelo estadístico, incluidos los contrastes de medias, los modelos ANOVA, o los modelos de regresión con observaciones correlacionadas o no correlacionadas. No obstante, suponemos como en el caso lineal, que las observaciones  $y = (y_1, \dots, y_n)^T$  no están correlacionadas. Y suponemos que se realizan los experimentos en las condiciones experimentales  $x_1, \dots, x_n$ . En este caso hay que acudir a la definición de matriz de información de Fisher. Tal matriz para un diseño  $\xi$  es

$$M(\xi, \theta) = \sum_{x \in \mathcal{X}} \xi(x) I(x, \theta),$$

donde

$$I(x, \theta) = E_h \left[ \left( \frac{\partial L(\theta)}{\partial \theta} \right) \left( \frac{\partial L(\theta)}{\partial \theta^T} \right) \right] = -E_h \left[ \left( \frac{\partial^2 L(\theta)}{\partial \theta^2} \right) \right].$$

Con  $L(\theta)$  se denota la función de verosimilitud de la variable respuesta  $y$ . El Teorema de Cramér–Rao nos proporciona la aproximación asintótica de la inversa de la matriz de información de Fisher a la matriz de covarianzas.

**Teorema 1.5 (Teorema de Cramér–Rao)**

Si  $E_\theta(\hat{\theta}) = \omega(\theta)$ , entonces

$$\Sigma_{\hat{\theta}} \geq n^{-1} \frac{\partial \omega(\theta)}{\partial \theta} M^{-1}(\xi, \theta) \left( \frac{\partial \omega(\theta)}{\partial \theta} \right)^T.$$

Además, el estimador de máxima verosimilitud alcanza esta cota asintóticamente. Si el estimador de máxima verosimilitud es centrado entonces  $\Sigma_{\hat{\theta}} \approx n^{-1} M^{-1}(\xi, \theta)$  asintóticamente.

Anotamos a continuación una versión del Teorema General de Equivalencia para modelos no lineales.

**Teorema 1.6 (GET para modelos no lineales)**

Para cualquier diseño  $\xi$ ,

$$\partial \Phi [M(\xi^*, \theta), M(\xi, \theta)] \geq 0.$$

Si  $\Phi$  es diferenciable,

$$\psi(x, \xi^*, \theta) \geq 0, x \in \chi$$

donde

$$\begin{aligned} \psi(x, \xi, \theta) &= \partial\Phi [M(\xi, \theta), I(x, \theta)] \\ &= \text{tr} \{ [I(x, \theta) - M(\xi, \theta)] \nabla\Phi [M(\xi, \theta)] \} \end{aligned}$$

### Demostración

Véase White (1973). ■

## 1.11. Críticas al diseño óptimo de experimentos

Hasta aquí se han definido algunas de las características importantes del diseño óptimo de experimentos. Es interesante terminar este capítulo mencionando algunas de las críticas que, durante su corta existencia de vida, le han ido surgiendo (López-Fidalgo, 2009).

En primer lugar se ha criticado la elección de un modelo a priori, sin la disponibilidad de datos que lo confirmen o desmientan. Citando a Box (1979) "los modelos, por supuesto, nunca son ciertos, pero afortunadamente sólo es necesario que sean útiles". Suele estar detrás de la elección la experiencia, la información previa, la intuición del experimentador, incluso en ocasiones los modelos se obtienen analíticamente. Por otro lado se han desarrollado criterios para discriminar entre modelos 'rivales' (López-Fidalgo, Tommasi y Trandafir, 2007 y 2008).

También se critica que, en los modelos no lineales, la matriz de información dependa de los parámetros del modelo haciendo necesario otorgarles unos valores nominales. Para minimizar el impacto de la elección de valores nominales se puede llevar a cabo un análisis de la sensibilidad respecto de dichos parámetros en los diseños  $\Phi$ -óptimos locales, explícitos o numéricos. Por otro lado existe la posibilidad de seleccionar diseños minimax, secuenciales o bayesianos para atajar este problema. Otro enfoque interesante son los diseños adaptativos (p. ej. Moler, Plo y San Miguel, 2006) en los que en cada paso los diseños utilizan las respuestas obtenidas previamente.

La selección de los criterios de optimización también es otro punto señalado como débil. Según las necesidades, se pueden considerar una composición de criterios para considerar diversos objetivos de forma conjunta (Cook y Wong, 1994). No obstante, los diseños óptimos no siempre difieren de unos criterios a otros, de hecho el teorema general de equivalencia establece como se ha mencionado la equivalencia de los diseños D- y G-óptimos. Es más, en algunos casos existen diseños universalmente óptimos (Harman, 2008). Otro problema con algunos criterios, es que la escala de los parámetros puede ser bastante diferente. La solución dada a este problema es considerar

criterios estandarizados de optimización basados en la eficiencia de cada parámetro, para una eficiencia final similar, o basados en el coeficiente de variación (López-Fidalgo y Rivas-López, 2007; López-Fidalgo, Rivas-López y Fernández-Garzón, 2007).

Un punto de gran controversia es la elección de diseños exactos, defendida por Box, o aproximados, defendida por Kiefer. Las propiedades de los diseños aproximados son muy convenientes a nivel teórico y computacional. Sin embargo, en la práctica deben redondearse con la consecuente pérdida de eficiencia. Por el contrario, los diseños exactos se adaptan a la práctica experimental pero son muy difíciles de obtener. Para tamaños de muestra pequeños tiene más sentido calcular diseños exactos, que con frecuencia no son tan difíciles de calcular (Pukelsheim y Rieder 1992, Imhof, López-Fidalgo y Wong 2001).

Tampoco faltan críticas sobre la dificultad de cálculo de diseños óptimos. Se han desarrollado algoritmos eficientes para la obtención de diseños óptimos, tanto exactos como aproximados. Cabe mencionar de entre todos el algoritmo de Wynn-Fedorov (Wynn, 1970; Fedorov, 1972) que básicamente añade puntos a un diseño inicial basándose en la información proporcionada por el teorema de equivalencia. También destacamos el algoritmo multiplicativo (Silvey, Titterington y Torsney, 1978; Torsney y Martín-Martín, 2009) que calcula los puntos del diseño transformando el espacio de diseño en proporciones definidas por los subintervalos entre puntos del diseño consecutivos. Éste último se ha utilizado durante esta tesis. La búsqueda de buenos algoritmos es un tema candente con mucho margen de desarrollo.

La última crítica de esta lista, no exhaustiva, es que los diseños óptimos con frecuencia exigen condiciones extremas. Por ejemplo, en un experimento clínico el diseño óptimo puede conducir a proporcionar la máxima dosis posible de un medicamento a un grupo de pacientes y ninguna dosis a otro grupo. Podría decirse que lo mejor es enemigo de lo bueno. No obstante, pueden proponerse otros diseños y proporcionar su eficiencia respecto del diseño óptimo, que sirve como medida de referencia. También pueden considerarse criterios restringidos, por ejemplo, a una dosis mínima de medicamento (p.ej. Ardanuy y López-Fidalgo, 1992).

## 1.12. Breve historia del diseño óptimo

Están cerca de cumplirse los 100 años de edad del Diseño Óptimo de Experimentos. Sus orígenes datan de 1918 cuando Smith publica un artículo donde establece la idea de diseño como distribución de probabilidad para una regresión polinomial. Establece un criterio que años más tarde Kiefer y Wolfowitz (1959) denominan G-optimización. Y, como en otras ramas de la ciencia, hay un salto temporal en el desarrollo de esta teoría hasta 1943 cuando Wald establece el criterio de maximización del determinante de la

matriz de información. Bautizado de nuevo por Kiefer y Wolfowitz en 1960 como D-optimización.

En unos pocos años se definen la mayoría de los criterios popularmente denominados alfabéticos, basados en funciones de la matriz de información siguiendo la propuesta de Kiefer. Así, Elfving (1952) emplea el criterio de la maximización de la traza de la matriz de información, A-optimización. El propio Elfving aborda el problema de la optimización de una combinación lineal de los parámetros introduciendo el criterio de c-optimización y proporcionando incluso un método gráfico para su cálculo. En 1955 Ehrenfeld establece la E-optimización (de Eigenvalues) que consiste en maximizar el mínimo autovalor de la matriz de información. Se va desarrollando así la llamada teoría convexa de diseños aproximados. La novedad estriba en considerar un diseño como medida de probabilidad. Algunos de sus seguidores son Atwood, Covey-Crump, Silvey, Fedorov, Karlin, Studden, Whittle, Wynn, etc.

En paralelo se desarrolla otra corriente de Diseño Óptimo de Experimentos liderada por Box, basada en la matriz de información para valorar la elección de los puntos de observación, sin emplear los citados criterios alfabéticos. Sus seguidores son N. R. Draper, J. S. Hunter, Lucas, Wilson y otros. Desde este punto de vista la generalización a funciones no polinómicas se hace problemática.

El siguiente paso importante es el establecimiento del Teorema General de Equivalencia. En 1958 Hoel hace de precursor comprobando en algunos casos que los criterios de Smith y de Wald, G- y D-optimización, dan los mismos resultados. Son Kiefer y Wolfowitz en 1959 los que demuestran el teorema y se erigen como grandes contribuidores al diseño óptimo de experimentos. Por su parte Box y Lucas (1959) aplicaron el criterio de D-optimización en modelos no lineales. Usando un argumento geométrico obtienen diseños de  $m$  puntos para modelos de  $m$  parámetros. Demuestran que el diseño D-óptimo maximiza el volumen del elipsoide de confianza de las estimaciones de los parámetros. Whittle (1973) generaliza el Teorema de Equivalencia para cualquier función criterio convexa, y al mismo tiempo White (1973) lo extiende a diseños para modelos no lineales. Kiefer (1974) da resultados de equivalencia para otros criterios.

Otro paso más es el desarrollo de métodos de construcción de diseño óptimos. De modo independiente Wynn (1970) y Fedorov (1972) son los primeros en desarrollar un método general para la construcción del diseño D-óptimo. Demuestran también que dicho algoritmo converge y dan un valioso procedimiento para calcular la matriz de información y su inversa en cada paso a partir de los cálculos hechos en el paso anterior. A modo de curiosidad, el libro de Fedorov es publicado en ruso en 1969 y en 1970 Wynn, que no sabía ruso, publica un artículo con el desarrollo del algoritmo de construcción de diseños D-óptimos. Cuando en 1972 se publica la traducción del libro de Fedorov hecha por Studden aparece en él reflejado básicamente el mismo

algoritmo. Hoy día se admite la producción independiente del algoritmo por los dos autores.

Box y Hunter en 1965 obtienen un algoritmo para la determinación del diseño D-óptimo en el modelo no lineal. Se trata esencialmente de una aplicación de la versión de los algoritmos sugeridos por Fedorov y Wynn a partir del Teorema de Equivalencia. Draper y Hunter (1967) discutieron el problema de seleccionar distribuciones de parámetros a priori con el objeto de obtener diseños para modelos no lineales. Atkinson y Hunter (1968) extendieron los resultados de Box y Lucas al caso en que el diseño toma más de  $m$  puntos. Box (1968a; 1968b; 1969; 1970) da algunos resultados adicionales para modelos no lineales. Por su parte Silvey y Titterington en 1973 dan una interpretación geométrica del diseño óptimo y plantean un algoritmo para obtener un diseño D-óptimo en el espacio dual. El propio Titterington (1976) ahondará en los aspectos geométricos del D-óptimo. Wu y Wynn (1978) dan condiciones generales para la convergencia de los algoritmos para la obtención del diseño óptimo. Hill (1980) demostró que si un modelo no es lineal en alguno de los parámetros, entonces el diseño D-óptimo no depende del valor de los parámetros en que es lineal. Abdelbasit y Plackett (1983) trabajan con modelos de regresión logística y obtienen diseños que maximizan la información sobre los parámetros en el modelo. Otras modificaciones de los métodos para obtener diseños óptimos en modelos no lineales se deben a Ford, Titterington y Kitsos (1989) y Khuri y Lee (1998).

Los libros de Fedorov (1972), Silvey (1980), Pázman (1986), Pukelsheim (1993), Schwabe (1996), que se centra en modelos multifactoriales, Fedorov y Hackl (1997), y el más reciente de Atkinson, Donev y Tobias (2007), actualización del libro de 1992 añadiendo rutinas de *SAS*<sup>®</sup>, son un buen compendio de los resultados más importantes obtenidos hasta esos momentos. En 1985 se publicó un libro recogiendo una colección de artículos de Kiefer sobre diseño óptimo de experimentos (Brown et al., 1985). Dicha colección es de un inestimable valor para los investigadores en esta materia. En España resaltar la aparición del libro de Rodríguez Torreblanca y Ortíz Rodríguez (1999), el que posiblemente sea el primer volumen en español dedicado íntegramente a Diseño Óptimo de Experimentos.

Entre las monografías más recientes dedicadas al diseño cabe destacar la aplicación de los métodos algebraicos al diseño de experimentos por parte de Pistone, Riccomagno y Wynn (2000) y el enfoque funcional por parte de Melas (2005) que está formado por un compendio de los trabajos del autor durante las dos últimas décadas. Uciński (2005) presenta métodos tanto clásicos como recientes de estimación óptima de parámetros en sistemas distribuidos dinámicos modelizados mediante ecuaciones diferenciales parciales. Una colección de trabajos aplicados a los campos de la Biología, Epidemiología, Medicina e Industria entre otros, ha sido recogida por Berger y Wong (2005). Estos mismos autores acercan las ideas y métodos del diseño óptimo a los investigadores de las áreas biomédica y social, limitando

su contenido matemático (Berger y Wong, 2009). En la misma línea, con un lenguaje ameno en forma de diálogos y basados en el estudio de casos, Goos y Jones (2011) describen aplicaciones prácticas del diseño de experimentos en Biotecnología, Industria, Química, etc. considerando experimentos de screening, diseños de superficies de respuesta, experimentos de mezclas, diseños en bloques, diseños split-plot, etc.

Aspectos más teóricos relacionados con modelos impredecibles, tamaños de muestra pequeños y conjuntos de datos perdidos/truncados e imputados se encuentran en Pronzato y Pázman (2013). A caballo entre la teoría y la práctica se encuentran los libros de Kitsos (2013) y de Fedorov y Leonov (2013). El primero describe el problema no lineal de diseño óptimo para modelos de cinética química mientras que el segundo se enfoca a los modelos no lineales habituales de la industria biofarmacéutica.

Entre los temas candentes se encuentran los diseños adaptativos, experimentos computacionales, observaciones correlacionadas, modelos mixtos con efectos aleatorios y experimentos con mezclas, por citar algunos. Hu y Rosenberger (2006) hacen una revisión de una amplia gama de diseños adaptativos, basados en las asignaciones y/o respuestas previas. En los ensayos clínicos estos diseños son fundamentales desde el punto de vista ético. Sobre experimentos computacionales se puede encontrar una revisión de la literatura en Levy y Steinberg (2010). En cuanto a los diseños con observaciones correlacionadas Zhigljavsky, Dette y Pepelyshev (2010) consideran un nuevo enfoque y Dette, Pepelyshev y Zhigljavsky (2014) hablan de diseños óptimos casi universales. Por su parte Amo-Salas, López-Fidalgo y Porcu (2013) obtienen diseños para procesos estocásticos cuya covarianza es función de la media. Müller y Stehlík (2009) consideran la aproximación de los resultados de experimentos simulados computacionalmente mediante observaciones correlacionadas. Sobre el diseño de experimentos en modelos mixtos caben destacar el libro de Demidenko (2004) y artículos como el de Dragalin y Fedorov (2006) o el de NGuyen, Bazzoli y Mentré (2012). El diseño en experimentos con mezclas (Cornell, 2002) es de gran complejidad práctica como señalan Coetzer y Haines (2013) en su artículo motivado por el proceso de gasificación del carbón.

Los algoritmos para el cálculo de diseños óptimos merecen mención aparte. Junto con el desarrollo de los algoritmos más populares y sencillos, Wynn-Fedorov, multiplicativo y de intercambio, se están utilizando algoritmos meta-heurísticos como los genéticos (Lin et al., 2014), simulated annealing (Woods, 2010; Wilmut y Zhou, 2011) o los basados en ‘enjambres de partículas’, entre otros (véase Mandal, Wong y Yu, 2014, para una recopilación). Sin embargo, estos novedosos algoritmos no están ganando mucha popularidad por su complejidad. Atkinson, Donev y Tobias (2007) describen diversas variantes de los algoritmos de intercambio, originalmente propuestos por Fedorov (1972). El libro de Pronzato y Pázman (2013) también proporciona un listado de algoritmos para el cálculo de diseños óptimos. Entre ellos describen el al-

goritmo branch-and-bound, propuesto por Welch (1982). Un ejemplo de su aplicación puede encontrarse en Ranjan, Bingham y Michailidis (2008).

La mayoría de estos algoritmos se deben ajustar manualmente para las aplicaciones prácticas. No obstante, existe una cierta variedad de software que implementa el cálculo de diseños óptimos o estos algoritmos. Por ejemplo, el cálculo de la matriz de información poblacional en los estudios de farmacocinética y farmacodinámica ha generado el desarrollo de los paquetes de software: PFIM (de Bazzoli, Retout y Mentré) en R, PkStaMP (de Leonov y Aliev), PopDes (de Ogungbenro), PopED (de Nyberg, Ueckert y Hooker) y WinPOPT/POPT (de Duffull), estos cuatro últimos programados en *Matlab*<sup>®</sup>. El algoritmo de Wynn-Fedorov está implementado en los cuatro primeros, mientras que los algoritmos de intercambio y simulated annealing lo implementan PopDes y POPT. A nivel general el libro de Atkinson, Donev y Tobias (2007) incluye códigos del software *SAS*<sup>®</sup> para el cálculo de diseños óptimos. Para *R* se ha desarrollado el paquete OPDOE (Rasch et al., 2011).

Por último, destacar la celebración periódica desde el año 1987, cada tres años, de una reunión de carácter internacional con el objetivo de reunir a investigadores de todo el mundo que trabajan en el campo del diseño óptimo. Los trabajos presentados en estas reuniones han sido publicados y forman un resumen de los avances de los investigadores más punteros en este ámbito. Una de estas reuniones, mODa 8, se celebró en el año 2007 en Almagro (España), bajo el auspicio de la Universidad de Castilla-la Mancha (López-Fidalgo, Rodríguez-Díaz y Torsney, 2007).

La más reciente, mODa 10, se celebró en 2013 en Lagów Lubuski (Polonia). Entre los temas tratados destacamos las aplicaciones a ensayos clínicos, estudios de búsqueda de dosis, diseños adaptativos, diseños para modelos no lineales de efectos mixtos, modelos compartimentales, experimentos de tiempo de vida y análisis de supervivencia. También se expusieron temas relacionados con observaciones correlacionadas, diseño para experimentos computacionales, discriminación entre modelos, regularización de criterios óptimos y otros surgidos de aplicaciones como los diseños para modelos de superficie de respuesta que involucran múltiples variables de procesos y mezclas. Una gran parte de las comunicaciones trataban de algoritmos, en el contexto de KL-optimización o criterios minimax, o desde las perspectivas funcional y bayesiana (Uciński, Atkinson y Patan, 2013). Una parte de esta tesis también fue presentada en una ponencia ante este público.

## Capítulo 2

# Modelos de ecuaciones simultáneas

Los modelos lineales de ecuaciones simultáneas son materia de estudio en los cursos de introducción a la Econometría. El paradigma de estos modelos es el modelo de oferta–demanda. Las ecuaciones del modelo relacionan la cantidad ofertada y la demandada de un bien con su precio. Cantidad y precio se determinan mutuamente y pueden estar correlacionadas. El objetivo es estimar simultáneamente, de aquí el nombre, los parámetros que relacionan las ecuaciones. Pero estos modelos no son exclusivos de la Economía. También se manejan en las ciencias sociales (p.ej. Mandal, Roe y Fein, 2014; Masten, 2014), la medicina (p.ej. Bentler y Stein, 1992; Kirkegaard et al., 2014) o la ingeniería (p.ej. Porra, 2002; García et al., 2014).

Este capítulo sólo pretende introducir la notación habitual de estos modelos así como señalar los principales métodos de estimación de los parámetros de interés del modelo. No se tendrán aquí en cuenta consideraciones de carácter temporal como es habitual en estos modelos. Para más detalles se pueden consultar Novales (2002), Gujarati (1997) ó Regúlez (2008) entre otros.

### 2.1. Modelo estructural de ecuaciones simultáneas

El modelo estructural consiste en especificar las ecuaciones de regresión que expresan la relación estructural de cada una de las variables endógenas con el resto de variables endógenas y exógenas o predeterminadas. Se supone una ecuación por cada variable endógena de interés. En ella pueden estar el resto, algunas o ninguna de las variables endógenas, así como todas, algunas o ninguna de las variables exógenas.

Supongamos un sistema completo de  $r$  ecuaciones estructurales estocásticas lineales de  $r$  variables dependientes o explicadas,  $Y_1, \dots, Y_r$  que se deter-

minan conjuntamente, y  $k$  variables explicativas predeterminadas  $X_1, \dots, X_k$ ,

$$\underbrace{Y_{1t}\gamma_{1i} + \dots + Y_{rt}\gamma_{ri}}_{\text{endógenas}} + \underbrace{X_{1t}\beta_{1i} + \dots + X_{kt}\beta_{ki}}_{\text{exógenas}} = U_{it}, \quad \begin{array}{l} i = 1, \dots, r \\ t = 1, \dots, n. \end{array}$$

donde  $n$  es el número de experimentos,  $\gamma_{li}$ ,  $l = 1, \dots, r$ , y  $\beta_{di}$ ,  $d = 1, \dots, k$ , son los parámetros estructurales que se pretenden estimar y  $U_{it}$  son los errores o perturbaciones estructurales aleatorias con media cero e independientes y homocedásticos entre observaciones, es decir, con matriz de covarianzas común,  $\text{cov}(U_{1t}, \dots, U_{rt}) = \Sigma$ , para cada observación. En forma matricial

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{rt} \end{pmatrix}^T \begin{pmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1r} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{r1} & \gamma_{r2} & \dots & \gamma_{rr} \end{pmatrix} + \begin{pmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{kt} \end{pmatrix}^T \begin{pmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1r} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \dots & \beta_{kr} \end{pmatrix} = \begin{pmatrix} U_{1t} \\ U_{2t} \\ \vdots \\ U_{rt} \end{pmatrix}$$

$$t = 1, \dots, T.$$

La forma más compacta de expresar el modelo estructural de ecuaciones simultáneas es

$$Y\Gamma + XB = U. \quad (2.1)$$

No obstante, la intuitiva notación anterior no resulta útil por no considerar satisfactoriamente las restricciones de nulidad de los parámetros de cada ecuación. Es decir, el hecho de que una ecuación contenga sólo algunas de las variables endógenas, porque la estructura de la ecuación indique que alguna de las variables endógenas del problema no determina la variable endógena respuesta considerada en dicha ecuación. Y lo mismo con las variables exógenas  $X_{it}$ , alguna puede no tener sentido para explicar una de las variables endógenas. En su lugar consideramos la notación siguiente. Consideramos cada una de las  $r$  ecuaciones de regresión para cada una de las variables endógenas, que ahora denotamos por  $y_1, \dots, y_r$ , poniendo dicha variable a un lado de la ecuación y el resto de variables endógenas y exógenas (sólo las que aparezcan en la ecuación) y el término de error/perturbación al otro lado. Además es habitual normalizar dicha ecuación, esto es, suponer su parámetro asociado igual a 1, lo que no supone ninguna pérdida de generalidad, basta dividir toda la ecuación por dicho parámetro. De este modo, cada ecuación de regresión se expresa como

$$y_{it} = Y_{it}\gamma_{it} + X_{it}\beta_{it} + u_{it} = Z_{it}\delta_{it} + u_{it}, \quad i = 1, \dots, r; \quad t = 1, \dots, n. \quad (2.2)$$

A modo de ilustración describimos el modelo de oferta–demanda. Las variables endógenas son la cantidad  $Q$  y el precio  $P$  de un bien, las variables

exógenas del modelo son la renta  $Y$  y el precio  $Z$  de un bien relacionado. El modelo estructural de oferta–demanda es:

$$Q_d = \gamma_1 P + \beta_0 + \beta_1 Y + \varepsilon_1, \quad (2.3)$$

$$Q_o = \gamma_2 P + \beta_3 + \beta_4 Z + \varepsilon_2, \quad (2.4)$$

junto con la condición de equilibrio del mercado  $Q = Q_d = Q_o$ . Este modelo expresa el hecho de que la cantidad demandada  $Q_d$  depende del precio del bien y la renta, mientras que la cantidad ofertada  $Q_o$  depende de los precios del bien y del bien relacionado. En forma matricial

$$\begin{pmatrix} Q \\ P \end{pmatrix}^T \begin{pmatrix} 1 & 1 \\ -\gamma_1 & -\gamma_2 \end{pmatrix} + \begin{pmatrix} 1 \\ Y \\ Z \end{pmatrix}^T \begin{pmatrix} -\beta_0 & -\beta_3 \\ -\beta_1 & 0 \\ 0 & -\beta_4 \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

## 2.2. Modelo reducido de ecuaciones simultáneas

El modelo reducido especifica cada una de las variables endógenas sólo en términos de las variables exógenas. Esto sólo es posible si  $\Gamma$  de (2.1) es invertible. En tal caso se podrán expresar los parámetros de la forma reducida como funciones, en general no lineales, de los parámetros de la forma estructural. El problema de la identificación se refiere al camino inverso, a obtener los parámetros de la forma estructural a partir de los parámetros estimados de la forma reducida. Se clasifican las ecuaciones estructurales en sobreidentificadas, exactamente identificadas y no identificadas, dependiendo de si se tiene más de una, una o ninguna combinación de valores de los parámetros de la forma estructural obtenidas las estimaciones de los parámetros de la forma reducida. Existen reglas para determinar la identificación de una ecuación, las conocidas como condición de orden y condición de rango. La última es una condición necesaria y suficiente, mientras que la primera sólo es necesaria.

Siguiendo la notación de (2.2) la del modelo reducido es

$$y_i = X\Pi_i + v_i, \quad i = 1, \dots, r, \quad (2.5)$$

donde  $X$  es la matriz  $n \times k$  de las  $k$  variables exógenas,  $\Pi_i$  es el vector  $k \times 1$  de coeficientes, y  $v_i$  es el vector  $n \times 1$  de errores de la forma reducida con matriz de covarianzas común,  $\text{cov}(v_1, \dots, v_r) = \Omega$ . En forma matricial,  $Y = X\Pi + V$ .

Siguiendo con el ejemplo del modelo oferta–demanda, resolviendo el sistema, por ejemplo, despejando el valor de  $P$  de la segunda ecuación (2.4) y sustituyendo en la primera (2.3) se consigue expresar la variable endógena  $Q$  sólo en términos de las variables exógenas  $(1, Y, Z)$ , su forma reducida,

$$Q = \begin{pmatrix} 1 \\ Y \\ Z \end{pmatrix}^T \begin{pmatrix} \frac{-\gamma_1\beta_3 + \gamma_2\beta_0}{\gamma_2 - \gamma_1} \\ \frac{\gamma_2\beta_1}{\gamma_2 - \gamma_1} \\ \frac{-\gamma_1\beta_4}{\gamma_2 - \gamma_1} \end{pmatrix} + v_1.$$

### 2.3. Estimación de los parámetros

La presencia en el modelo estructural de varias variables explicativas endógenas estocásticas conlleva que el estimador de mínimos cuadrados ordinarios, OLSE, de cada ecuación sea insesgado. De aquí el interés en el modelo reducido. El estimador OLS de los parámetros de la forma reducida  $\hat{\Pi} = (X^T X)^{-1} X^T Y$  es insesgado. Además dicho estimador es equivalente a estimar por OLS cada una de las ecuaciones de la forma reducida por separado.

Existen numerosos métodos de estimación de los parámetros estructurales del modelo (2.1). Por un lado los métodos que estiman por separado cada una de las ecuaciones del sistema, también conocidos como métodos uniecuacionales o de información limitada. Entre ellos cabe destacar los métodos de mínimos cuadrados indirectos, mínimos cuadrados en dos etapas, 2SLS, o máxima verosimilitud con información limitada. Por otro lado los que estiman conjunta y simultáneamente todos los parámetros del sistema de ecuaciones identificadas exactamente o sobreidentificadas, denominados de información completa. Los dos métodos de estimación de este tipo más habituales son el de estimación mínimo cuadrática en tres etapas, 3SLS, y el de máxima verosimilitud con información completa, FIML. Que son los utilizados durante el desarrollo de esta tesis.

Zellner y Theil (1962) introdujeron el método 3SLS que estima los coeficientes estructurales en tres etapas, generalizando el método de 2SLS en el sentido de tomar en consideración las correlaciones contemporáneas de los términos de error. En la primera etapa se estima la matriz de momentos de las perturbaciones de la forma reducida. En la segunda se estiman los parámetros de cada una de las ecuaciones estructurales, individualmente. Utilizando la matriz de covarianzas estimada por mínimos cuadrados en dos etapas se estiman simultáneamente en la tercera etapa todos los parámetros del sistema completo de ecuaciones mediante mínimos cuadrados generalizados. El estimador 3SLS coincide con el 2SLS cuando no existen correlaciones entre ecuaciones o si todas las ecuaciones están exactamente identificadas.

Rothenberg y Leenders (1964) proporcionan una expresión explícita para la matriz de covarianzas asintótica de los estimadores FIML de los parámetros estructurales que es equivalente a la de los estimadores 3SLS, cuando los errores se supone que siguen una distribución normal. Este resultado es de gran ayuda pues el método de máxima verosimilitud es costoso, computacionalmente hablando, como muchos autores han señalado y como veremos en la Sección 3.3.

Tanto el estimador 3SLS como el FIML tienen buenas propiedades. Son consistentes cuando las ecuaciones están identificadas. Y son asintóticamente eficientes.

## Capítulo 3

# Diseños óptimos en modelos de ecuaciones simultáneas

En este capítulo se van a describir los tres problemas considerados y las soluciones aportadas durante el desarrollo de esta tesis. Fruto de este trabajo son los dos artículos ya publicados y el enviado para su publicación que se pueden encontrar en los Anexos.

### 3.1. Diseños experimentales en modelos triangulares de ecuaciones simultáneas

Se considera el problema de diseño óptimo en un modelo de regresión lineal donde alguna de las variables explicativas no puede ser controlada por el experimentador. Además se considera que no es posible conocer el valor de las variables no controlables antes de realizar el experimento. López-Fidalgo y Garcet-Rodríguez (2004) desarrollan la teoría de diseños óptimos condicionalmente restringidos, que se denotan con CR, para afrontar este problema. Se basan en el caso clínico de Varela et al. (2001) que estudian la utilidad de la oximetría en un ejercicio preoperatorio estandarizado, pedaleo en bicicleta estática, para predecir la morbilidad cardiopulmonar después de la resección pulmonar. El tiempo de pedaleo es la variable controlable y la desaturación en sangre, medida en el dedo por oximetría durante dicho ejercicio, es la variable no controlable con valores obviamente desconocidos antes de realizar el experimento. López-Fidalgo y Garcet-Rodríguez (2004) obtienen diseños óptimos considerando una única ecuación de regresión con dos variables explicativas, una controlable/diseñable,  $x$ , y otra no controlable,  $z$ , de la que se desconoce su valor antes de realizar el experimento. Suponen también que se dispone de información a priori sobre la variable no controlable a través de un diseño condicional conocido  $\tilde{\xi}_{z|x}(\cdot|x)$ , esto es una distribución de probabilidad de los posibles valores de  $z$  conocidos los valores de  $x$ .

La novedad que se aporta en esta tesis es afrontar este problema considerando un modelo de ecuaciones simultáneas, que en adelante denotaremos por SE. La variable no controlable se considera primero como explicativa, exógena en la terminología de ecuaciones simultáneas, en la ecuación de regresión para la variable respuesta  $y$ . Y se considera como respuesta, endógena, en una segunda ecuación, conformando un modelo triangular de dos ecuaciones simultáneas. Es decir, la misma variable juega el papel de endógena y exógena, en ecuaciones distintas. En ambas ecuaciones se considera la variable controlable  $x$ , que será diseñada utilizando el criterio de D-optimización (véase la Sección 1.7).

La diferencia entre ambos planteamientos estriba en las mayores o menores suposiciones sobre la variable no controlable. El primero considera sólo la primera ecuación suponiendo totalmente conocida la distribución de la variable no controlable  $z$ , la simultáneamente exógena y endógena, mediante un diseño condicional. Por su parte el enfoque SE supone para dicha variable una distribución de probabilidad parcialmente conocida, que debe ajustarse. Por esto es razonable pensar que el enfoque SE es más general que el CR. Además el planteamiento SE permite capturar la interacción entre las variables endógenas (Conlisk, 1979). Las comparaciones entre ambos, que veremos más adelante, no son con el objetivo de determinar cual de los dos planteamientos es mejor. Esta decisión, como se viene señalando, dependerá de la información que el experimentador tenga a priori sobre el problema.

En la literatura también se ha considerado el modelo de regresión con una variable explicativa no controlable conocida a priori, lo que conduce a diseños marginalmente restringidos, o una variable no controlable conocida y otra desconocida, que lleva a plantear diseños marginal y condicionalmente restringidos. Mientras que en esta sección se obtienen explícitamente los diseños óptimos, en situaciones más complejas y/o reales se puede utilizar el algoritmo multiplicativo para construir diseños marginal y condicionalmente restringidos (Martín–Martín, Torsney y López–Fidalgo, 2007).

Los modelos triangulares de ecuaciones simultáneas son un caso particular de éstos. No por ello son sencillos de manejar. Además, en la literatura se han encontrado significativos ejemplos en los que se utilizan dichos modelos. Por ejemplo Newey, Powell y Vella (1999) investigan la relación no lineal empírica entre el ratio del salario por hora y el número de horas anuales trabajadas. Una estructura triangular en la que se podría aplicar lo desarrollado en esta sección es el modelo dosis–respuesta considerado por Biswas y López–Fidalgo (2013) donde la toxicidad es exógena en la ecuación de la eficacia, pues se espera que la toxicidad afecte a la eficacia, y a su vez es endógena en una segunda ecuación en la que la eficacia no interviene, pues no se espera que influya en la toxicidad.

**Matrices de información CR y SE**

Se consideran la distribución condicional  $y|z$ , determinada por la *pdf*  $g(y|z; x, \theta_1)$  con media  $\eta_{12}(z; x, \theta_1)$  y  $\text{var}(y|z) = \sigma_{12}^2(z; x, \theta_1)$ , y la distribución marginal de  $z$ , determinada por la *pdf*  $h(z; x, \theta_2)$ . A partir de la función de verosimilitud se obtienen las matrices de información, FIMs, para un diseño conjunto  $\xi(x, z)$  tanto para el planteamiento CR como para el SE, siendo  $\theta^T = (\theta_1^T, \theta_2^T)$

$$M_{CR}(\theta_1, \xi_{CR}) = \sum_{x,z} \xi_{CR}(x) \tilde{\xi}_{z|x}(z|x) E_{y|z;x} \left[ -\frac{\partial^2 \log g(y|z; x, \theta_1)}{\partial \theta_1 \partial \theta_1^T} \right]$$

$$M_{SE}(\theta, \xi_{SE}) = \sum_x \xi_{SE}(x) E_z \left[ E_{y|z} \left( -\frac{\partial^2 [\log g(y|z; x, \theta_1) + \log h(z; x, \theta_2)]}{\partial \theta \partial \theta^T} \right) \right].$$

Para llegar a estas expresiones se ha tenido en cuenta que  $\xi_{CR}(x, z) = \xi_{CR}(x) \tilde{\xi}_{z|x}(z|x)$  y que la distribución conjunta de  $y$  y  $z$  es  $f(y, z; x, \theta) = g(y|z; x, \theta_1)h(z; x, \theta_2)$ .

Y se tiene el siguiente resultado.

**Teorema 3.1** *Si  $M_{CR}$  es la FIM de un diseño condicionalmente restringido  $\xi_{CR}$ ,  $M_{SE}$  es la FIM del diseño de ecuaciones simultáneas correspondiente  $\xi_{SE}$  y  $\theta_1$  y  $\theta_2$  son los parámetros de la primera y segunda ecuación, respectivamente, entonces:*

1. *Si  $h(z; x, \theta_2) = \tilde{\xi}_{z|x}(\cdot|x)$  y se conoce totalmente de antemano:  $M_{SE} = M_{CR}$  y por tanto  $\xi_{SE}^* = \xi_{CR}^*$  para cualquier criterio de optimalidad.*
2. *Si  $\theta_1$  y  $\theta_2$  son desconocidas para el enfoque SE y no tienen componentes en común, entonces*

$$M_{SE}(\theta, \xi_{SE}) = \begin{pmatrix} M_{CR}(\theta_1, \xi_{SE}) & 0 \\ 0 & M_{II}(\theta_2, \xi_{SE}) \end{pmatrix},$$

*donde  $M_{II}$  es la FIM de la segunda ecuación. Esta estructura diagonal por bloques es consecuencia de la estructura triangular del modelo SE.*

La demostración de este Teorema se encuentra en la Sección 2.1 del anexo A.1 (véase Teorema 2).

**Diseños D-óptimos para espacios de diseño de dos puntos**

A partir de las matrices de información y suponiendo las condiciones del punto 2. del Teorema 3.1 se obtienen en un primer paso los diseños D-óptimos para espacios de diseño de dos puntos de soporte,

$$\xi^*(x) = \left\{ \begin{array}{cc} 0 & 1 \\ 1 - p^* & p^* \end{array} \right\}, \quad 0 \leq p^* = \arg \min |M^{-1}(\theta, \xi)| \leq 1.$$

Se considera el modelo de regresión lineal

$$E(y|z; x) = \eta_{12}(z; x, \theta_1) = \alpha x + \beta z,$$

con posibles valores para la variable no controlable  $z \in \mathcal{Z} = \{0, 1\}$  y con la siguiente información a priori,

$$\tilde{\xi}_{z|x}(\cdot|0) = \begin{Bmatrix} 0 & 1 \\ 1 - r_0 & r_0 \end{Bmatrix}, \quad \tilde{\xi}_{z|x}(\cdot|1) = \begin{Bmatrix} 0 & 1 \\ 1 - r_1 & r_1 \end{Bmatrix}.$$

Esto indica que el  $100r_0\%$  de los valores de  $z$  serán 1 cuando  $x$  toma el valor 0 siendo el  $100r_1\%$  cuando  $x$  toma el valor 1. Por tanto,  $z$  sigue una distribución de Bernoulli para cada valor de  $x$  con media  $r_x$  conocida.

El correspondiente modelo SE considerado es,

$$\begin{cases} E(y|z; x) = \eta_{12}(z; x, \theta_1) = \alpha x + \beta z, \\ E(z|x) = \eta_2(x; \theta_2) = r_x, \end{cases}$$

donde la variable no controlable  $z$  se modeliza mediante la probabilidad de  $z = 1$  y se supone la distribución normal para la variable respuesta  $y$ .

Los correspondientes pesos D-óptimos son:

$$p_{CR}^* = \xi_{CR}^*(1) = \begin{cases} \frac{r_0}{2[r_0 - r_1(1 - r_1)]} & r_0 > 2r_1(1 - r_1), \\ 1 & r_0 \leq 2r_1(1 - r_1), \end{cases}$$

$$p_{SE}^* = \frac{6r_0 + 3(r_1 - 1)r_1 - \sqrt{4r_0^2 + 4r_0(r_1 - 1)r_1 + 9(r_1 - 1)^2r_1^2}}{8(r_0 + (r_1 - 1)r_1)}.$$

Se observa aquí claramente la dependencia de los diseños respecto de los valores nominales de los parámetros. Esto se debe a la no linealidad de los parámetros de ambos modelos. Por ello se ha realizado el correspondiente análisis de sensibilidad, como se comenta más adelante.

Además ambos pesos son simétricos respecto al parámetro  $r_1$ . Las diferencias entre sus valores se pueden apreciar en la Figura 3.1. En el planteamiento CR se obtienen dos tipos de diseños D-óptimos, unipuntuales concentrados en  $x = 1$  para los valores de  $(r_0, r_1)$  de la zona sombreada (a la izquierda de la curva  $r_0 = 2r_1(1 - r_1)$ ) y con dos puntos de soporte. En el planteamiento SE sólo se obtienen diseños D-óptimos bipuntuales, con valores de  $p_{SE}^* = \xi_{SE}^*(1)$  acotados entre 0,5 y 0,75, para cualquier par de valores  $(r_0, r_1)$ .

Al calcular las eficiencias relativas (véase la Sección 3.3 del anexo A.1) de los diseños de ambos planteamientos se tienen mejores eficiencias relativas de los diseños SE respecto de los CR que viceversa. Para la región sombreada de la Figura 3.1 (izquierda) los diseños CR son ineficientes desde el punto de vista del planteamiento SE.

Por otro lado, se ha realizado el correspondiente análisis de sensibilidad mediante el cálculo de la eficiencia relativa del diseño con unos parámetros

### 3.1 Diseños en modelos triangulares de ecuaciones simultáneas 31

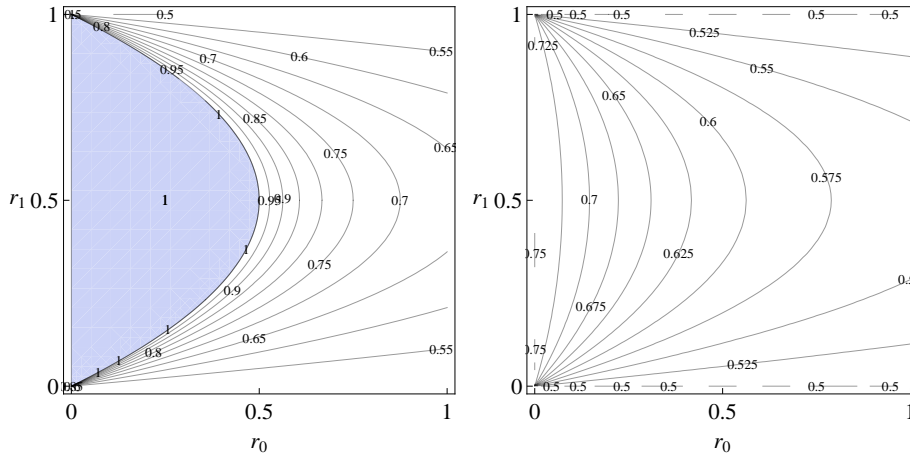


Figura 3.1: Valores de  $p^*$  para los diseños CR D-óptimos (izquierda) y SE D-óptimos (derecha) para  $\chi = \{0, 1\}$  y  $(r_0, r_1) \in [0, 1] \times [0, 1]$ .

nominales dados frente al diseño con los parámetros reales (véase (3) en la Sección 3.3 del anexo A.1). Dicha eficiencia indica la bondad del diseño de parámetros nominales frente al diseño con los parámetros reales. Del análisis se desprende que los diseños SE son muy robustos. Las eficiencias quedan por encima del 90 % en la mayoría de los casos. De hecho, ante el desconocimiento de los valores reales de los parámetros, la mejor opción es tomar como valores nominales el punto medio  $r_0 = r_1 = 1/2$  para el que la eficiencia relativa respecto al diseño con cualesquiera parámetros reales no baja del 97 %. Sin embargo los diseños CR no son tan robustos, puesto que en uno de los escenarios considerados las eficiencias bajan del 50 %. En el anexo A.1 se pueden encontrar más detalles y figuras que ilustran estos resultados.

#### Diseños D-óptimos para espacios de diseño continuos

Es natural extender los espacios de diseño de dos puntos considerando anteriormente a espacios de diseño continuos con extremos en dichos puntos. La primera pregunta planteada es si los diseños obtenidos anteriormente siguen siendo óptimos cuando se considera el espacio continuo. Para comprobarlo se ha hecho uso del Teorema General de Equivalencia, GET (véase la Sección 1.8). Concretamente de la adaptación del GET al planteamiento CR (López-Fidalgo y Garcet-Rodríguez, 2004, Teorema 4) y de una versión más general tomada de Pázman (1986), Proposición IV.26, para el enfoque SE (véanse las secciones 4.1 y 4.2 del anexo A.1).

Suponiendo una función lineal en  $x$  para  $r_x$  (probabilidad de que  $z = 1$  dado el valor de  $x$ ) los diseños CR D-óptimos para el espacio de diseño  $\chi = \{0, 1\}$  lo son también para  $\chi = [0, 1]$  excepto en la región superior

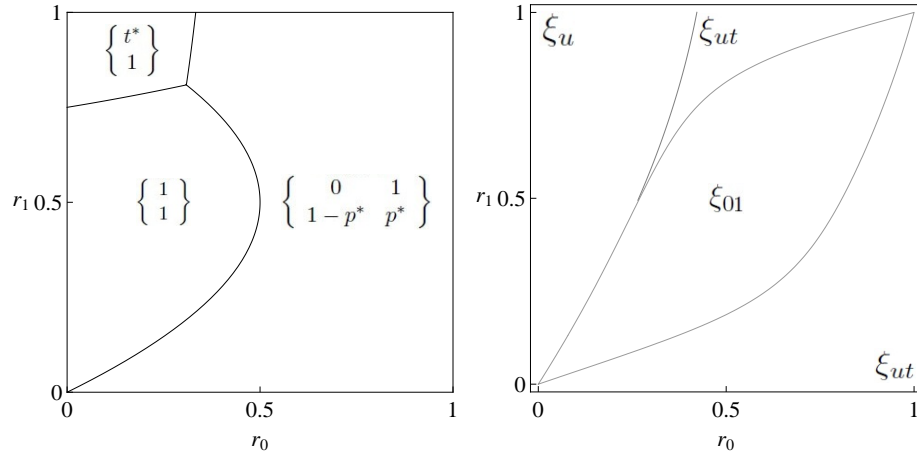


Figura 3.2: Diseños D-óptimos para  $\chi = [0, 1]$  para los modelos CR (izquierda) y SE (derecha).

izquierda como se aprecia en la Figura 3.2 (izquierda). En dicha región se obtiene un diseño CR D-óptimo concentrado en  $0 < t^* < 1$ . Mientras que los diseños SE D-óptimos en  $\chi = \{0, 1\}$ , denotados con  $\xi_{01}$  en la Figura 3.2 (derecha), sólo son D-óptimos para  $\chi = [0, 1]$  en las cercanías de la diagonal principal. Para el resto de posibles valores de  $r_0$  y  $r_1$  los diseños SE D-óptimos obtenidos son de uno,  $\xi_u$ , o dos puntos de soporte,  $\xi_{ut}$ , siendo  $u, t \in (0, 1)$ .

Respecto a la robustez, a pesar de cambiar los diseños CR y SE, los diseños SE siguen siendo más robustos que los CR. Para los diseños CR D-óptimos con  $\chi = [0, 1]$ , el análisis de sensibilidad señala que considerar valores nominales de la región superior izquierda no es la mejor elección si se desconocen los parámetros reales. Y en el caso de los diseños SE D-óptimos para el espacio de diseño continuo la peor opción es considerar el punto medio como valores nominales, justo al contrario de lo observado cuando el espacio de diseño es de dos puntos.

Estos resultados se basan en suponer una función lineal para  $r_x$ . Si se considera una función no lineal como  $r_x = \sin(10x) + [r_1 - r_0 - \sin(10)]x + r_0$  los diseños CR D-óptimos para  $\chi = \{0, 1\}$  no son óptimos para  $\chi = [0, 1]$ . Un caso de especial interés es considerar un función logística para  $r_x$  de la forma  $1/(1 + e^{-(\lambda + \delta x)}) \in (0, 1)$  con parámetros  $\theta_2 = (\lambda, \delta)^T \in \mathbb{R}^2$ , desconocidos y estimables. Como en el caso lineal, los diseños SE D-óptimos para  $\chi = \{0, 1\}$  son óptimos para  $\chi = [0, 1]$  en un entorno de la diagonal principal. Dicho entorno es más amplio que en el caso lineal, pero se va estrechando conforme aumenta  $T$  si se considera el intervalo  $[0, T]$  (véase la Figura 3.3). El correspondiente análisis de sensibilidad señala como regiones a evitar para los parámetros nominales la superior izquierda y la inferior izquierda, por

### 3.1 Diseños en modelos triangulares de ecuaciones simultáneas 33

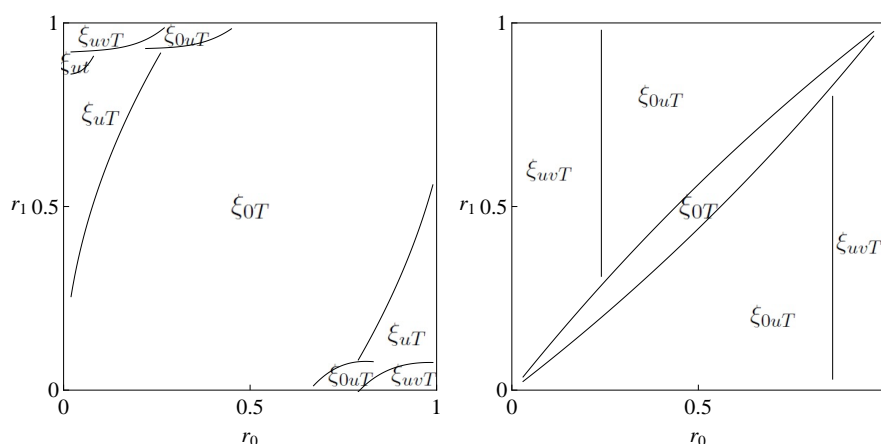


Figura 3.3: Diseños SE D-óptimos para  $\chi = [0, T]$  cuando  $T = 1$  (izquierda) y  $T = 100$  (derecha).

simetría.

#### El problema de la identificación

Por último, es inherente al planteamiento SE el problema de la identificación (véase Sección 2.2). En el caso aquí considerado las dos ecuaciones están identificadas, es decir, pueden estimarse sin ambigüedades todos los parámetros, para cualquier  $r_x$ . Sin embargo, si se añade un término independiente a la primera ecuación y  $r_x$  es lineal, dicho término independiente y el parámetro que acompaña a  $z$  no estarían identificados, aunque una función no lineal de  $r_x$  podría ayudar a identificarlos.

**Observación:** Poskitt y Skeels (2008) indican que los diseños para modelos de ecuaciones simultáneas dependen de las dos formulaciones posibles, estructural (SES) o reducida (RFS). Los dos modelos conducen a diferentes funciones de verosimilitud y por tanto diferentes matrices de información y diseños óptimos. El modelo de ecuaciones simultáneas considerado en esta sección está formulado en forma estructural. Sin embargo en la obtención de los diseños, al considerar la función de verosimilitud, manejamos la forma reducida, a la que llegamos simplemente sustituyendo la segunda ecuación en la primera. En todo momento consideramos los parámetros estructurales y no los de la forma reducida. Por ello es importante asegurar que sean estimables/identificables, como lo son en este caso gracias a su estructura singular y a la linealidad de la primera ecuación.

A modo de comprobación se han obtenido diseños D-óptimos y c-óptimos

para los modelo SES y RFS definidos en Poskitt y Skeels (2008).

$$\begin{cases} y &= Y\beta + u, \\ Y &= \Pi_{2a} + Z\Pi_{2b} + V, \end{cases} \quad \begin{pmatrix} u \\ V \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \quad (SES)$$

$$\begin{cases} y &= \Pi_{2a}\beta + Z\Pi_{2b}\beta + \nu, \\ Y &= \Pi_{2a} + Z\Pi_{2b} + V, \end{cases} \quad \begin{pmatrix} \nu \\ V \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho' \\ \rho' & 1 \end{pmatrix} \right). \quad (RFS)$$

En este caso la variable diseñable es  $Z$ . Se considera un espacio de diseño de dos puntos  $\mathcal{Z} = \{0, 1\}$ . Como ya se ha visto, los pesos óptimos dependen de los valores nominales. Para  $(\beta, \Pi_{2a}, \Pi_{2b}, \rho = \rho') = (3, 4, 1, 0.8)$  donde el valor de  $\beta$  sólo es necesario al manejar el modelo RFS cuyos diseños no dependen de  $\rho'$ :

	D-opt		$c_\beta$ -opt		$c_{\Pi_{2a}}$ -opt		$c_{\Pi_{2b}}$ -opt		$c_{\rho, \rho'}$ -opt	
	SES	RFS	SES	RFS	SES	RFS	SES	RFS	SES	RFS
$\rho$ conocida	.547	.553	1		0	.528	.50092	.5089	-	
$\rho$ estimable	.548	.553					.50097		[0,1]	

Como se puede apreciar los pesos de los diseños D- y c-óptimos obtenidos, aunque distintos, son similares para los dos modelos. Además haciendo uso del Teorema General de Equivalencia los diseños D-,  $c_{\Pi_{2a}}$ - (sólo para el modelo RFS) y  $c_{\Pi_{2b}}$ -óptimos para el espacio de diseño  $\mathcal{Z} = \{0, 1\}$  también son óptimos para el espacio de diseño  $\mathcal{Z} = [0, 1]$ .

Aunque existe dependencia de los parámetros nominales, los pesos de los diseños D-óptimos están acotados así como la D-eficiencia, que nos da idea del riesgo que podríamos correr al seleccionar equivocadamente los valores nominales de los parámetros. Concretamente, para los modelos SES y RFS tanto si la correlación es conocida como si es estimable, el peso del diseño D-óptimo en el espacio de diseño  $\mathcal{Z} = \{0, 1\}$ , para cualquier valor de los parámetros  $\beta$ ,  $\Pi_{2a}$ ,  $\Pi_{2b}$  y  $|\rho| < 1$  ó  $|\rho'| < 1$  que definen el modelo, está acotado:

$$1/3 \leq p_D^* \leq 2/3.$$

La cota mínima de D-eficiencia para el modelo SES o el modelo RFS es

- cuando la correlación es conocida:  $\text{mín D-eff}_{\theta^*}(\xi_{\theta^{(0)}}) = \frac{1}{2^{1/3}},$
- y cuando la correlación es estimable:  $\text{mín D-eff}_{\theta^*}(\xi_{\theta^{(0)}}) = \frac{1}{2^{1/4}},$

para cualesquiera valores reales  $\theta^*$  y nominales  $\theta^{(0)}$  de los parámetros.

### Pseudo-Paradoja diseño exacto-aproximado

Se ha escrito (Sección 1.4) que si  $n$  es suficientemente grande podría aproximarse un diseño aproximado a uno exacto. Vamos a presentar una

### 3.1 Diseños en modelos triangulares de ecuaciones simultáneas 35

pseudo-paradoja de esta aproximación que no se incluyó en el artículo publicado. Aprovechando el modelo triangular de ecuaciones simultáneas en un sencillo espacio de diseño obtendremos un diseño aproximado que siempre diferirá en su aplicación práctica del correspondiente diseño exacto.

Supóngase el modelo SE siguiente:

$$\left. \begin{aligned} y &= \beta z + u \\ z &= \delta x + v \end{aligned} \right\},$$

donde  $y$  es una variable aleatoria respuesta que depende de la variable aleatoria  $z$ , que a su vez depende de la variable controlable  $x$ ;  $\beta$  y  $\delta$  son parámetros desconocidos y  $u$  y  $v$  son los términos de error experimental que, en línea con la teoría general de modelos lineales, suponemos que siguen una distribución normal de media 0 y varianza conocida.

Sustituyendo la segunda ecuación en la primera el sistema sólo depende de  $x$ ,

$$\left. \begin{aligned} y &= \beta\delta x + \beta v + u \\ z &= \delta x + v, \end{aligned} \right\} \quad \left. \begin{aligned} y &= \pi_1 x + \epsilon \\ z &= \pi_2 x + \epsilon'. \end{aligned} \right\}$$

Los estimadores del modelo y su matriz de covarianzas se pueden obtener por el método de estimación de mínimos cuadrados ordinarios:

$$\begin{pmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \end{pmatrix} = \begin{pmatrix} \frac{\overline{xy}}{\overline{x^2}} \\ \frac{\overline{xz}}{\overline{x^2}} \end{pmatrix} \quad y \quad Cov \begin{pmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \end{pmatrix} = \frac{\overline{x^2}}{x^2} \begin{pmatrix} \sigma_y^2 & \sigma_{yz} \\ \sigma_{yz} & \sigma_z^2 \end{pmatrix}.$$

En esta matriz de covarianzas sólo se desconocen los valores de  $x$ , la variable diseñable. Vamos a obtener diseños D-óptimos. Esto es, se minimiza el determinante de la matriz de covarianzas anterior:  $\overline{x^2}/\overline{x^2}$ .

Tomamos  $\mathcal{X} = \{0, 1\}$  como espacio de diseño de la variable  $x$ . Para un tamaño  $n$  cualquiera se tendría:

$$\min_{x_1, \dots, x_n} \left( \frac{\overline{x^2}}{\overline{x^2}} \right) = \min_{x_1, \dots, x_n} \left( \frac{\left( \frac{\sum_{i=1}^n x_i}{n} \right)^2}{\frac{\sum_{i=1}^n x_i^2}{n}} \right) = \min_{x_1, \dots, x_n} \left( \frac{1}{n} + \frac{2 \sum_{i,j=1; j>i}^n x_i x_j}{n \sum_{i=1}^n (x_i^2)} \right).$$

El mínimo se alcanzará cuando el numerador del último término sea mínimo y el denominador máximo. Tomando  $n - 1$  puntos del diseño el valor 0 y el otro el valor 1, el numerador será nulo pero el denominador no, alcanzándose el mínimo. Se llega así al diseño D-óptimo exacto de tamaño  $n$ :

$$\xi_n = \left\{ \begin{array}{cc} 0 & 1 \\ n-1 & 1 \end{array} \right\}.$$

Para encontrar el diseño D-óptimo aproximado consideramos el siguiente diseño:

$$\xi = \left\{ \begin{array}{cc} 0 & 1 \\ 1-p & p \end{array} \right\}.$$

donde  $p$  caracteriza la medida discreta de probabilidad, siendo  $p \in [0, 1]$ . Se han de buscar los valores de  $x$  que minimizan  $\bar{x}^2/\bar{x}^2$ . Teniendo en cuenta que  $x_i$  y  $x_i^2$  toman el valor 0 con probabilidad  $1 - p$  y el valor 1 con probabilidad  $p$ , la expresión a minimizar queda:

$$\min_{x_1, \dots, x_n} \left( \frac{\bar{x}^2}{\bar{x}^2} \right) = \min_{x_1, \dots, x_n} \left( \frac{\left( \frac{\sum_{i=1}^n x_i}{n} \right)^2}{\frac{\sum_{i=1}^n x_i^2}{n}} \right) = \min_p \left( \frac{\left( \frac{np}{n} \right)^2}{\frac{np}{n}} \right) = \min_p p$$

El mínimo se alcanza para  $p = 0$  siendo el diseño D-óptimo aproximado:

$$\xi = \left\{ \begin{array}{c} 0 \\ 1 \end{array} \right\},$$

es decir, todo el diseño se concentra en el punto 0.

Se ha encontrado un diseño D-óptimo aproximado que carga todo el diseño en un punto, y el correspondiente diseño D-óptimo exacto que siempre tiene dos puntos de soporte, independientemente del tamaño del diseño experimental. Es decir, ambos diseños siempre diferirán en su aplicación práctica. Aunque ambos diseños se aproximan al aumentar el tamaño del diseño:

$$\xi_n = \left\{ \begin{array}{cc} 0 & 1 \\ \frac{n-1}{n} & \frac{1}{n} \end{array} \right\} \xrightarrow{n \rightarrow \infty} \xi = \left\{ \begin{array}{c} 0 \\ 1 \end{array} \right\}.$$

### 3.2. Criterio robusto de composición integral

Al considerar en la Sección 3.1 anterior dos modelos distintos para afrontar el problema surge de manera natural la comparación entre ambos. Como ya se ha dicho, la diferencia entre los planteamientos es sobre todo conceptual y depende de la información disponible sobre el problema real, por lo que no se considera un problema de rivalidad de modelos. No obstante, se ha investigado una nueva clase de criterios para obtener diseños robustos respecto de ambos planteamientos. Este desarrollo surge de las conversaciones mantenidas con el profesor Milan Stehlík durante la estancia del doctorando en la Johannes Kepler University de Linz (Austria). El profesor Stehlík junto con Elena Bukina venían trabajando en la relación entre un criterio  $D_\alpha$ -óptimo compuesto para estimar tanto los parámetros de tendencia como los de la estructura de covarianzas y el criterio integral del error de predicción cuadrático medio, IMSPE. Fruto de esta colaboración es la introducción de una nueva clase de criterios de composición integral.

De manera general se puede decir que si existen dos o más enfoques para un problema dado, por ejemplo, dos modelos rivales o dos partes diferentes de un modelo, la composición integral de las respectivas matrices de información, FIMs, puede ser útil para construir diseños que tengan una información

similar para los diferentes modelos, esto es, diseños robustos. La comparación de FIMs de varios modelos se ha investigado desde diversos puntos de vista en la literatura estadística (véase p.ej. Ahmadi y Arghami, 2003; Hofmann, 2004; o Alshunnar, Raqab y Kundu, 2012). Sin embargo los criterios compuestos clásicos pueden no ser adecuados para obtener diseños robustos como veremos en el ejemplo del siguiente párrafo. A partir de las FIMs de ambos modelos o ambas partes se optimiza un criterio de composición integral respecto de una función de densidad de una familia paramétrica de distribuciones de probabilidad.

En modelos con observaciones correlacionadas, donde la tendencia y la estructura de covarianzas tienen parámetros diferentes la FIM es una matriz diagonal por bloques, como ocurre en el planteamiento SE considerado en la Sección 3.1 anterior (véase el Teorema 1.1). Poner ambas matrices al mismo nivel puede ser inapropiado. El método propuesto en esta Sección considera este tipo de situaciones produciendo diseños robustos para estimar ambas partes del modelo. El criterio compuesto clásico puede no prestar atención a una parte del modelo. Así ocurre al considerar  $Y(x_i) = \varepsilon(x_i)$ ,  $x_i \in [0, 1]$  con los errores correlacionados mediante una función de covarianzas, o bien exponencial,  $\exp(-\theta d)$ , o bien lineal,  $1 - \theta d$ , donde  $d = |x_i - x_j|$ . En este caso, considerando diseños D-óptimos exactos de dos puntos,  $(x, x + d)$ , para el parámetro de interés  $\theta > 0$ , se llega a que  $d = 0$ , es decir el diseño D-óptimo colapsa, tanto considerando correlación exponencial como el criterio compuesto  $\theta \log M_{exp}(\theta) + (1 - \theta) \log M_{lin}(\theta)$ . Sin embargo el caso de correlación lineal conduce a  $d^* = \min\{2/\theta, 1\}$ . Para más detalles véase la sección 1.3 del Anexo A.3.

### Composición integral

Consideremos dos modelos, que pueden diferir en su representación estadística. Sean  $M_1(\theta)$  y  $M_2(\theta)$  las respectivas FIMs con  $\theta \in \Theta$ . Sea  $u : \Theta \rightarrow R$  una función de un conjunto  $D$ , eventualmente una familia paramétrica de funciones, p.ej. de densidad. Consideremos el criterio de composición integral

$$Lu = \int_{\Theta} \phi(\theta, M_1(\theta), M_2(\theta))u(\theta)d\theta, \quad (3.1)$$

donde  $\phi$  es un kernel determinado, que mide la discrepancia entre las dos FIMs. Hay que señalar que el vector de parámetros  $\theta$  incluye los parámetros del modelo y otros parámetros o constantes relacionadas con el procedimiento, p. ej. nuevos parámetros que definen el kernel.

El objetivo es encontrar una función  $u^*$  apropiada, p. ej. que maximice o minimice  $Lu$ , o que resuelva la ecuación integral de Fredholm  $Lu = \psi$ , donde  $\psi$  es un criterio dado. Para esta función se construirá un diseño, que proporcionará un diseño óptimo robusto adecuado. Obsérvese que la optimización de  $Lu$  puede diferir de la optimización de su kernel  $\phi$ .

Ejemplos típicos de kernel,  $\phi$ , son

1.  $\phi_1(\theta, M_1(\theta), M_2(\theta)) = |M_1(\theta)|^{\theta_3/m_1} |M_2(\theta)|^{(1-\theta_3)/m_2}$ .
2.  $\phi_2(\theta, M_1(\theta), M_2(\theta)) = \frac{\theta_3}{m_1} \log |M_1(\theta)| + \frac{1-\theta_3}{m_2} \log |M_2(\theta)|$ .
3.  $\phi_{2^*}(\theta, M_1(\theta), M_2(\theta)) = \theta_3 \log |M_1(\theta)| + (1 - \theta_3) \log |M_2(\theta)|$ .
4.  $\phi_3(\theta, M_1(\theta), M_2(\theta)) = \theta_3 \left( \frac{|M_1(\theta)|}{|M_1^*(\theta)|} \right)^{1/m_1} + (1 - \theta_3) \left( \frac{|M_2(\theta)|}{|M_2^*(\theta)|} \right)^{1/m_2}$ , donde  $|M_i^*(\theta)|$  corresponde al valor máximo de  $|M_i(\theta)|$ ,  $i = 1, 2$ .
5.  $\phi_{3^*}(\theta, M_1(\theta), M_2(\theta)) = \theta_3 \left( \frac{|M_1(\theta)|}{|M_1^*(\theta)|} \right) + (1 - \theta_3) \left( \frac{|M_2(\theta)|}{|M_2^*(\theta)|} \right)$ .

Con  $|\cdot|$  denotamos tanto el determinante como el valor absoluto,  $\theta^T = (\theta_1, \theta_2, \theta_3)$ ,  $0 \leq \theta_3 \leq 1$  y  $m_i$  es la dimensión del subvector  $\theta_i$  de  $\theta$ ,  $i = 1, 2$ .

Tras encontrar la función apropiada  $u$  se busca el diseño óptimo bien minimizando o maximizando  $Lu(\xi)$ , o bien construyendo el diseño utilizando  $u$ . La función  $u$  puede fijarse desde el principio si existe algún interés o conocimiento sobre los parámetros del problema. Un ejemplo trivial es la medida de Dirac, que conduce a optimizar localmente el kernel. Otras distribuciones típicas son la uniforme o la familia Beta. Algunos kernels producen el mismo espectro de diseños óptimos (véase el Lema 1, Sección 1.2 anexo A.2).

Si la función  $u(\theta)$  no depende de los parámetros del modelo(s) se tiene el criterio compuesto clásico (véase el Lema 2 Sección 1.2 anexo A.2). Los criterios para discriminar entre modelos rivales, p. ej. T- o KL-optimización (Atkinson y Fedorov, 1975a y 1975b, López-Fidalgo, Tommasi y Trandafir, 2007) se pueden considerar un caso particular de la clase de criterios definidos en este trabajo teniendo en cuenta algunas generalizaciones no consideradas aquí.

Como complemento a la definición se estudian algunas propiedades. Se proporciona, por medio de operadores integrales de Fredholm, la base teórica para la existencia de un criterio específico, relacionado con el criterio  $IMSPE(x|\xi) = \int_{X^n} \sigma^{-2} MSPE(x|\xi) d\xi(x)$ . Este último criterio busca predecir con precisión en toda la región de diseño  $\chi = X^n$  minimizando la llamada varianza Kriging  $\text{Var}[\hat{Y}(x|\xi)] = E[(\hat{Y}(x|\xi) - Y(x))^2] \equiv MSPE(x|\xi)$  (Error de Predicción Cuadrático Medio) donde  $\hat{Y}(x|\xi)$  denota el mejor predictor lineal insesgado de  $Y(x)$  basado en un diseño particular  $\xi$ . El criterio IMSPE se ha utilizado para construir diseños óptimos (véase Sacks, Schiller y Welch, 1989, ó Crary, 2002).

También se justifica la necesidad de regularizar la ecuación de Fredholm. Además se ilustra la metodología con varios experimentos numéricos (véase Anexo A.2).

### Criterio de composición integral para diseños condicionalmente restringidos y de ecuaciones simultáneas

En esta memoria nos centraremos en la aplicación del criterio de composición integral para comparar las FIMs de los planteamientos CR y SE de la Sección 3.1 anterior.

Recordemos que el enfoque CR considera un diseño conjunto de las variables explicativas  $(x, z)$  teniendo en cuenta que  $z$  no puede diseñarse y sólo se conocen sus valores después de realizar el experimento. Se fija el diseño condicional  $\xi_{z|x}(\cdot|x)$ , que indica el resultado de  $z$  que se espera dependiendo del valor  $x$ . Se supone completamente conocido, basándose por ejemplo en datos históricos o en la experiencia del experimentador. Se obtiene por esta vía la matriz de información  $M_{CR}$ . Desde el planteamiento SE la variable no controlable  $z$  se modeliza a través de una nueva distribución de probabilidad, que debe ajustarse con la información disponible, y se obtiene la matriz de información  $M_{SE}$ . Cabe señalar que no se define una estructura de correlación entre distintas observaciones como se ha comentado anteriormente sino entre las dos variables respuestas del modelo SE.

Consideramos para ambos enfoques el espacio de diseño  $\chi = \{0, 1\}$ , el espacio de posibles valores de  $z$ ,  $\mathcal{Z} = \{0, 1\}$  y suponemos que la respuesta sigue una distribución normal con media

$$E(y|x, z) = \eta_1(y|x, z; \theta_1) = a + bx + cz,$$

donde le hemos añadido el término independiente a la considerada en la Sección 3.1.

La variable no controlable  $z$  se diseña (CR)/modeliza (SE) mediante la probabilidad de  $z = 1$  dado el valor de  $x$ , que denotamos por  $r_x$ . Por simplicidad, se supone conocida la varianza,  $\sigma_y^2 = 1$ , y  $r_0 = 0$ . Mediante la experimentación se estimaría  $r_1$ . Así, el vector de parámetros es  $\theta^T = (\theta_1^T, \theta_2)$  con  $\theta_1^T = (a, b, c) \in \mathbb{R}^3$  y  $\theta_2 = r_1 \in [0, 1]$ .

Dado que el modelo no es lineal la FIM depende del parámetro  $\theta_2 = r_1$ . Calculando las FIMs como en la Sección 3.1

$$\begin{aligned} M_{SE}(r_1, \xi) &= \begin{pmatrix} 1 & p & pr_1 & 0 \\ p & p & pr_1 & 0 \\ pr_1 & pr_1 & pr_1 & 0 \\ 0 & 0 & 0 & \frac{p}{r_1(1-r_1)} \end{pmatrix} \\ &= \left( \begin{array}{c|c} M_{CR}(r_1, \xi) & 0 \\ \hline 0 & M_{II}(r_1, \xi) \end{array} \right), \end{aligned}$$

donde los subíndices de las matrices indican las dos ecuaciones diferentes consideradas.

Optimizaremos el criterio (3.1) con respecto a la densidad de una familia paramétrica, a saber,  $u_{\alpha, \beta}(r_1)$  con distribución Beta de parámetros  $\alpha$  y  $\beta$ .

Seguiremos el procedimiento descrito anteriormente considerando el siguiente kernel, basado en las eficiencias,

$$\phi_3(\theta, M_{CR}(\theta), M_{II}(\theta)) = r_1 \left( \frac{|M_{CR}(\theta)|}{|M_{CR}^*(\theta)|} \right)^{1/3} + (1 - r_1) \frac{|M_{II}(\theta)|}{|M_{II}^*(\theta)|},$$

donde  $M_{CR}^*(\theta)$  y  $M_{II}^*(\theta)$  son las matrices de información que maximizan los correspondientes determinantes. En particular,  $|M_{CR}^*(\theta)| = 4r_1(1 - r_1)/27$  para  $p_{CR}^* = 2/3$  y  $|M_{II}^*(\theta)| = 1/[r_1(1 - r_1)]$  para  $p_{II}^* = 1$ . Por lo tanto,

$$\phi_3(\theta, M_{CR}(\theta), M_{II}(\theta)) = r_1 \left[ \frac{27p^2(1 - p)}{4} \right]^{1/3} + (1 - r_1)p$$

y

$$\begin{aligned} Lu(p) &= \int_0^1 \phi_3(r_1, M_{CR}(r_1), M_{II}(r_1)) u_{\alpha, \beta}(r_1) dr_1 \\ &= \frac{\beta p + \frac{3}{2^{-2/3}} \alpha [p^2(1 - p)]^{1/3}}{\alpha + \beta}. \end{aligned}$$

Supongamos que la probabilidad de  $z = 1$  (p.ej. desaturación en sangre) en el grupo experimental (p. ej. haciendo un ejercicio físico,  $x = 1$ ) es 0,4 en media con una varianza de 0,04. Esto implica que  $\alpha = 2$  y  $\beta = 3$ . Entonces el valor de  $p$  que maximiza  $Lu(p)$  es  $p^* = \frac{1}{3} \left( 2 - \frac{1}{\sqrt[3]{1 + \sqrt{2}}} + \sqrt[3]{1 + \sqrt{2}} \right) = 0,8654$ . Las eficiencias relativas son

$$\begin{aligned} \text{Eff}_{II}(p_{CR}^*) &= \frac{2/[3r_1(1 - r_1)]}{1/[r_1(1 - r_1)]} = 2/3 = 0,6667, \\ \text{Eff}_{II}(p^*) &= \frac{0,8654/[r_1(1 - r_1)]}{1/[r_1(1 - r_1)]} = 0,8654, \\ \text{Eff}_{CR}(p_{II}^*) &= 0, \\ \text{Eff}_{CR}(p^*) &= \left( \frac{(1 - 0,8654)0,8654^2 r_1(1 - r_1)}{4r_1(1 - r_1)/27} \right)^{1/3} = 0,879551, \end{aligned}$$

por lo que el diseño de composición integral es bastante robusto para ambas matrices.

### 3.3. Diseños óptimos con restricciones de costes en modelos de ecuaciones simultáneas

Llegamos al tercer problema considerado, que consiste en añadir restricciones de costes a un problema de diseño. Los diseños óptimos de experimentos con restricciones han sido tratados ampliamente en la literatura (véase

p.ej. Cook y Fedorov, 1995). Aquí consideramos restricciones vía costes de experimentación. Se definirá una función criterio compuesta por un criterio estadístico y una función de los costes. Mediante un coeficiente escalar se buscará el adecuado compromiso entre la información proporcionada por las observaciones, determinada por el criterio estadístico, y los costes de tomar dichas observaciones (Pronzato, 2010). Pero la principal novedad es considerar este planteamiento para modelos de ecuaciones simultáneas.

Como ya se ha comentado en la introducción, el primero que consideró el problema de diseño de experimentos para modelos de ecuaciones simultáneas fue Conlisk (1979). Propuso una función objetivo para el diseño basada en la matriz asintótica de covarianzas de los estimadores de los parámetros estructurales. Rothenberg y Leenders (1964) obtuvieron una expresión explícita para dicha matriz utilizando el método de máxima verosimilitud con información completa, FIML, a la que también se llega utilizando el método de mínimos cuadrados en tres etapas, 3SLS (Zellner y Theil, 1962) si se supone normalidad para la distribución de las perturbaciones (véase la Sección 2.3). Con dicha expresión se simplifican los cálculos, computacionalmente costosos para el método FIML como muchos autores han señalado, pero se basa en la forma reducida del modelo, como en la estimación 3SLS. Al utilizar la forma reducida se tiene no linealidad en los parámetros lo que implica que el criterio de diseño dependerá de los valores nominales que se le asignen a los parámetros a estimar y por tanto desconocidos del modelo.

La diferencia entre el desarrollo realizado en la Sección 3.1 y lo que se describe a continuación estriba en el cálculo de la FIM. Aquí se utilizará la inversa de la citada expresión de Rothenberg y Leenders (1964) mientras que en la Sección 3.1 se obtuvo directamente por el método de máxima verosimilitud. Además ahora se están considerando restricciones de costes.

### Problema de diseño con restricciones en un modelo de ecuaciones simultáneas

Consideramos el modelo estructural de ecuaciones simultáneas (2.2) descrito en la Sección 2.1. La expresión de la matriz asintótica de covarianzas de los estimadores 3SLS o FIML de los parámetros estructurales  $\delta$  suponiendo una distribución normal para las perturbaciones es

$$\text{plim}(\text{cov } \hat{\delta}) = \left[ -n \lim_{n \rightarrow \infty} E \left( \frac{\partial^2 L^*}{\partial \delta \partial \delta^T} \right) \right]^{-1} = [H^T (\Sigma^{-1} \otimes (X^T X)) H]^{-1}, \quad (3.2)$$

donde plim indica el límite en probabilidad,  $L^*$  es la denominada función de verosimilitud ‘concentrada’ (véase Rothenberg y Leenders, 1964),  $\otimes$  repre-

senta el producto de Kronecker de matrices y

$$H = \begin{pmatrix} H_1 & & 0 \\ & \ddots & \\ 0 & & H_r \end{pmatrix}, \quad (3.3)$$

donde  $H_i = (\Pi_i, J_i)$ ,  $\Pi_i$  viene definida por (2.5) y  $J_i$  se define como  $X_i = XJ_i$ , la matriz de vectores unitarios que seleccionan de  $X$  las variables exógenas incluidas en la  $i$ -ésima ecuación.

Para ilustrar el procedimiento que introducimos en esta sección definiremos un problema de diseño L-óptimo, con restricciones en los costes (para la definición de L-optimización véase la Sección 1.7). No obstante, el procedimiento se puede aplicar a cualquier criterio de optimización como por ejemplo D-, A-, c-optimización... Para definir el problema de diseño es necesario concretar el espacio de diseño  $\mathcal{X}$ , la función objetivo y las restricciones de costes, si las hay. Aunque el espacio de diseño sea continuo, la búsqueda del diseño óptimo se puede restringir a diseño con un número finito de puntos en su soporte por el corolario del Teorema de Caratheodory (véase la Sección 1.6). La función objetivo queda determinada por las matrices  $M$  y  $L$ , en el caso concreto de L-optimización. Suponemos que al experimentador le interesa la suma de las estimaciones de las varianzas de los errores para predecir las variables endógenas  $y_i$ ,  $i = 1, \dots, r$ , las definidas en el lado derecho de cada ecuación estructural, evaluadas en cada uno de los puntos del espacio finito de diseño  $\mathcal{X} = \{x_1, \dots, x_m\}$ . Por último se considera la restricción de costes  $\sum_{j=1}^m c_j n_j \leq C$ , donde  $c_j$  es el coste de tomar una observación en  $x_j$ ,  $n_j$  es el número de veces que  $x_j$  aparece en el diseño final y  $C$  es el presupuesto total del experimento.

Por tanto, el problema de diseño L-óptimo puede expresarse como

$$\begin{aligned} \min_{n_1, \dots, n_m} \operatorname{tr} & \left\{ \left[ H^T (I_r \otimes \sum_j x_j^T x_j) H \right] \left[ H^T \left( \Sigma^{-1} \otimes \sum_j n_j x_j^T x_j \right) H \right]^{-1} \right\} \\ \text{s.t.} & \sum_j n_j c_j \leq C \text{ y } n_1 \geq 0, \dots, n_m \geq 0, \end{aligned} \quad (3.4)$$

donde  $I_r$  es la matriz identidad de orden  $r$ .

Conlisk (1979) señala que, aunque los  $n_j$  deben ser enteros, diseño exacto, el procedimiento habitual es aproximarlos resolviendo el problema de programación no lineal continua y redondeando los valores de  $n_j$ .

### Un procedimiento iterativo eficiente

A la vista del problema anterior y dadas las complicaciones de cálculo del problema de programación no lineal continua se ha desarrollado un procedimiento iterativo eficiente para resolver el problema de diseño con

restricciones de costes. Se basa en el algoritmo multiplicativo, aplicable a diseños aproximados que pueden utilizarse aquí dado que  $X^T X$  de (3.2) se puede expresar como  $\sum_j n_j x_j^T x_j$  en (3.4), esto es, como suma ponderada de las matrices de información de los puntos del soporte.

El algoritmo multiplicativo (p.ej. Torsney y Martín–Martín, 2009) actualiza simultáneamente en cada paso todos los pesos de los puntos de soporte del diseño,  $\xi(x_j)$ , hasta proporcionar un diseño óptimo. Gracias a su uso se consigue una de las ventajas del procedimiento que describimos a continuación, la reducción en el tiempo de cómputo, en torno a 10 veces más rápido como veremos más adelante.

Cook y Wong (1994) demostraron que plantear un problema de diseño con restricciones y un problema de diseño óptimo compuesto son equivalentes. Así, el problema (3.4) es equivalente a obtener el diseño óptimo aproximado  $\xi^*$  que minimiza la función criterio compuesta y convexa

$$\Phi_\beta(\xi) = \Phi(\xi) - \beta B(\xi), \quad (3.5)$$

donde  $\xi$  es un diseño aproximado,  $\Phi$  es una función criterio de optimización,  $B(x_j)$  es una función apropiada de beneficios, basada en la función de costes conocida  $c_j$ ,  $B(\xi) = \sum_j \xi(x_j) B(x_j)$ , que es lineal en  $\xi$ , y  $\beta > 0$  es un parámetro ajustable. Obsérvese el signo menos que precede a  $\beta$ , para obtener el máximo beneficio/mínimo coste.

**Observación:** Utilizar directamente la función de costes en la definición del criterio compuesto no permitía el uso del algoritmo multiplicativo. De aquí que se haya considerado una función de beneficios apropiada. La linealidad en  $\xi$  de esta función, junto con la convexidad del criterio  $\Phi$ , hacen que la función criterio compuesta sea convexa:

$$\begin{aligned} \Phi_\beta[(1-\lambda)\xi + \lambda\xi'] &= \Phi[(1-\lambda)\xi + \lambda\xi'] - \beta B[(1-\lambda)\xi + \lambda\xi'] \leq \\ &\leq (1-\lambda)[\Phi(\xi) - \beta B(\xi)] + \lambda[\Phi(\xi') - \beta B(\xi')]. \end{aligned}$$

El Teorema General de Equivalencia, GET (Kiefer y Wolfowitz, 1959) indica que  $\xi^*$  es  $\Phi_\beta$ -óptimo sí y sólo si

$$\partial\Phi_\beta(\xi, 1_{x_j}) = \text{tr}\{\nabla\Phi(\xi)[M(1_{x_j}) - M(\xi)]\} - \beta[B(1_{x_j}) - B(\xi)] \geq 0, \quad j = 1, \dots, m,$$

donde  $\partial\Phi(\xi, \xi')$  es la derivada direccional de  $\Phi$  en  $\xi$  en la dirección de  $\xi'$  (véase la Sección 1.8) y  $1_{x_j}$  es el diseño concentrado en  $x_j$ , i.e. con peso 1 en  $x_j$ .

Basándose en el GET se define el algoritmo multiplicativo. Para cada valor de  $\beta$ , en el paso  $s+1$  los pesos de los puntos de soporte del diseño son actualizados mediante la expresión

$$\xi^{(s+1)}(x_j) = \xi^{(s)}(x_j) \frac{\beta B(1_{x_j}) - \text{tr}\{\nabla\Phi(\xi^{(s)})M(1_{x_j})\}}{\beta B(\xi^{(s)}) - \text{tr}\{\nabla\Phi(\xi^{(s)})M(\xi^{(s)})\}}. \quad (3.6)$$

Con esta formula  $\sum_j \xi^{(s+1)}(x_j) = 1$ .

Para determinar el diseño óptimo se establece una regla de parada basada en la cota inferior de eficiencia de cada nuevo diseño  $\xi^{(s+1)}$ . Dicha cota se obtiene aplicando el GET y depende del signo de  $\Phi_\beta(\xi^{(s+1)})$ ,

- Si  $\Phi_\beta(\xi^{(s+1)}) > 0$  entonces  $\text{eff}_{\Phi_\beta}[M(\xi^{(s+1)})] \geq 1 + K$ ,
- Si  $\Phi_\beta(\xi^{(s+1)}) < 0$  entonces  $\text{eff}_{\Phi_\beta}[M(\xi^{(s+1)})] \geq 1/(1 + K)$ ,

donde  $K = \min_{x_j} \partial \Phi_\beta(\xi^{(s+1)}, 1_{x_j}) / \Phi_\beta(\xi^{(s+1)})$ .

El procedimiento para obtener el diseño  $\Phi_\beta$ -óptimo  $\xi^*$  para el criterio (3.5) es:

1. Para cada valor de  $\beta$  se obtiene numéricamente un diseño óptimo aproximado,

$$\xi^{(\beta)} = \arg \min_{\xi} \Phi_\beta(\xi) = \arg \min_{\xi} [\Phi(\xi) - \beta B(\xi)],$$

utilizando el algoritmo multiplicativo (3.6).

2. Se obtiene  $\beta^* = \arg \min_{\beta} \Phi(n^{(\beta)} \xi^{(\beta)})$ , donde  $n^{(\beta)} = C/C(\xi^{(\beta)})$  es el número total de experimentos para cada  $\beta$  siendo  $C(\xi^{(\beta)}) = \sum_j \xi^{(\beta)}(x_j) c_j$ .
3. Por último  $n^{(\beta^*)} \xi^{(\beta^*)}(x_j)$  debe redondearse a números enteros,  $n_j^*$ , de modo que  $\sum_j n_j^* \approx n^{(\beta^*)}$  y  $\sum n_j^* c_j \approx C$ .

Obsérvese que el procedimiento puede aplicarse a cualquier problema de diseño con restricciones de costes.

### Un ejemplo ilustrativo

Una compañía de estaciones de servicio considera el modelo estructural de demanda de sus dos productos, gasolina y reparaciones,  $Q_g$  y  $Q_r$ , y decide experimentar con variaciones controladas en sus precios  $P_g$  y  $P_r$  (Conlisk, 1979). Considerando sólo estas variables explicativas y tres valores normalizados posibles (0,93, 1, 1,07) para cada una se tiene un diseño factorial de  $3^2$  condiciones experimentales:

$$\begin{aligned} \mathcal{X} = \{ & x_1 = (0,93, 0,93), & x_2 = (0,93, 1), & x_3 = (0,93, 1,07), \\ & x_4 = (1, 0,93), & x_5 = (1, 1), & x_6 = (1, 1,07), \\ & x_7 = (1,07, 0,93), & x_8 = (1,07, 1), & x_9 = (1,07, 1,07) \}. \end{aligned}$$

Para resolver el correspondiente problema de diseño experimental (3.4) se deben proporcionar valores nominales para  $\Sigma$ , la matriz de covarianzas de las perturbaciones, y  $H$ , que depende de  $\Pi$  via (3.3) y que a su vez depende de los parámetros estructurales  $\alpha_i$  y  $\beta_i$  de forma no lineal. De aquí

proviene la no linealidad del modelo. Sean los valores nominales dados por las ecuaciones:

$$\begin{cases} Q_g = 1,38 + 0,1Q_r - 0,5P_g + u_1, \\ Q_r = 0,39 + 0,8Q_g - 0,2P_r + u_2, \end{cases} \quad \Sigma = \begin{pmatrix} 0,1 & 0,06 \\ 0,06 & 0,1 \end{pmatrix}.$$

Resolviendo el sistema de ecuaciones estructurales para  $Q_g$  y  $Q_r$  se obtiene la forma reducida del modelo y con ella los valores nominales de  $\Pi_1$  y  $\Pi_2$ . Por lo tanto las matrices  $H_i$  toman valores

$$H_1 = \begin{pmatrix} 1,6239 & 1 & 0 \\ -0,4348 & 0 & 1 \\ -0,2174 & 0 & 0 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 1,5424 & 1 & 0 \\ -0,5438 & 0 & 0 \\ -0,0217 & 0 & 1 \end{pmatrix}.$$

Finalmente, se define el coste de realizar un experimento en la condición experimental (estación de servicio)  $j$ -ésima mediante

$$c_j = 5 + 3673(P_g^{(j)} - 1)^2 + 2857(P_r^{(j)} - 1)^2 + 1633(P_g^{(j)} - 1)(P_r^{(j)} - 1),$$

en términos del porcentaje de beneficios perdidos por la estación de servicio durante el experimento. Los nueve costes unitarios redondeados son  $c_j = \{45, 23, 29, 19, 5, 19, 29, 23, 45\}$  donde se aprecia simetría. El mínimo coste unitario corresponde a la condición experimental ‘central’, en la que se aplican los precios estándar  $P_g = P_r = 1$ . Los mayores costes son los de las condiciones experimentales ‘inicial’ y ‘final’ ( $P_g = P_r = 0,93$  y  $P_g = P_r = 1,07$ , respectivamente) lo que indica mayores pérdidas de beneficios por desviarse de los precios estándar. Se fija el presupuesto del experimento en  $C = 6500$  lo que equivale a la pérdida total de beneficios de 65 estaciones.

Con estas especificaciones se ha buscado el diseño óptimo resolviendo el problema (3.4). No se ha encontrado un único mínimo global. Los diseños óptimos obtenidos dependen del punto de inicio de búsqueda. Para ilustrarlo se detallan tres casos,

- Caso 1. Diseño inicial de búsqueda,  $n_j = 1$ ,  $j = 1, \dots, m$ .
- Caso 2. Diseño inicial de búsqueda,  $n_j = n_j^{\text{máx}}/2$ , donde  $n_j^{\text{máx}} = \arg \text{máx}\{n_j | n_j c_j = 6500\}$ ,  $j = 1, \dots, m$ .
- Caso 3. Diseño inicial de búsqueda,  $n_j = n_j^{\text{máx}} - 1$ .

Los tres diseños exactos (redondeados) óptimos correspondientes se presentan en la Tabla 3.1. Conforme aumentan los valores iniciales, los  $n_j$  óptimos de los puntos de soporte impares decrecen, aumentando los de los puntos pares. En todos los casos el punto de diseño ‘central’, el más barato, recibe el mayor número de experimentos.

Para los tres casos el valor de la función objetivo (3.4) de los diseños aproximados que son solución del problema es 0,020973, de aquí que los

	$P_g$ 0,93			1			1,07		
	$P_r$ 0,93	1	1,07	0,93	1	1,07	0,93	1	1,07
Caso 1	9	45	19	55	77	55	19	45	9
Caso 2	13	37	23	47	92	47	23	37	13
Caso 3	19	25	29	36	116	36	29	26	19
P	37	0	39	0	180	0	39	0	37

Tabla 3.1: Diseños exactos óptimos para estimar los parámetros estructurales del ejemplo numérico. Los Casos 1, 2 y 3 son soluciones directas del problema (3.4). P indica la solución obtenida con el procedimiento establecido en esta sección.

tres casos son mínimos globales. Dicha función objetivo toma los valores 0,0211127, 0,021129 y 0,0209315 para los casos 1, 2 y 3 respectivamente si se consideran los diseños exactos óptimos de la Tabla 3.1. Obsérvese que el último valor es menor que el de los diseños aproximados.

Si se consideran otras variables controlables,  $D_g$  y  $D_r$ , que reflejan si se ofrecen o no sellos de comercio con la venta de gasolina o reparaciones respectivamente, el espacio de diseño  $\mathcal{X}$  pasa a ser de dimensión 4,  $(P_g, P_r, D_g, D_r)$ . Asignando los valores 0 ó 1 a  $D_g$  y  $D_r$  se tienen  $3^2 \times 2^2 = 36$  condiciones experimentales posibles,

$$\mathcal{X} = \{x_1 = (0,93, 0,93, 0, 0), x_2 = (0,93, 0,93, 0, 1), x_3 = (0,93, 0,93, 1, 0), \\ x_4 = (0,93, 0,93, 1, 1), x_5 = (0,93, 1, 0, 0), \dots, x_{36} = (1,07, 1,07, 1, 1)\}.$$

Se añade al coste  $c_j$  el término  $5(1 - D_g^{(j)})$  para penalizar la ausencia de sellos de comercio en la venta de gasolina. Los costes unitarios se presentan en la Tabla 3.2 donde de nuevo se aprecia simetría. Las condiciones experimentales centrales siguen siendo las más baratas y las más caras resultan las de los extremos ( $P_g$  y  $P_r \neq 1$ ).

	$P_r$	0,93		1		1,07	
		$D_r$ 0	1	0	1	0	1
$P_g$	$D_g$						
0.93	0	50	50	28	28	34	34
	1	45	45	23	23	29	29
1	0	24	24	10	10	24	24
	1	19	19	5	5	19	19
1.07	0	34	34	28	28	50	50
	1	29	29	23	23	45	45

Tabla 3.2: Costes unitarios redondeados.

		$P_r$ 0,93 1 1,07								$P_r$ 0,93 1 1,07					
		$D_r$		0 1		0 1		0 1		$D_r$		0 1		0 1	
$P_g$	$D_g$														
0.93	0	3	1	17	7	6	11	0.93	0	4	16	0	0	17	6
	1	1	0	2	0	1	4		1	0	0	0	0	0	0
1	0	17	2	49	57	1	6	1	0	0	0	0	0	0	0
	1	8	0	137	136	1	18		1	0	0	165	166	0	0
1.07	0	3	1	0	1	0	0	1.07	0	10	27	0	0	26	9
	1	11	5	5	15	0	3		1	0	0	0	0	0	0

Conlisk (1979)
 Procedimiento de esta Sección

Tabla 3.3: Diseños L-óptimos para estimar los parámetros estructurales.

Suponemos valores nominales de 0,04 y 0,02 para los parámetros que acompañan a las variables  $D_g$  y  $D_r$  respectivamente. El diseño óptimo estructural para este ejemplo, proporcionado por Conlisk (1979) se presenta en la Tabla 3.3. Este diseño exacto alcanza los 529 experimentos con un coste total de 6471. Conlisk no proporciona el diseño aproximado que lleva a dicho diseño exacto. En nuestras comprobaciones hemos obtenido diseños aproximados similares al diseño de Conlisk con un valor inferior (mejor) de la función objetivo (3.4). Esto refuerza la idea de que el redondeo es un tema crucial como veremos más adelante.

Aplicamos ahora el procedimiento descrito más arriba en esta sección. Como ya se ha comentado, se considera como función criterio la de L-optimización. Consideramos la función beneficio  $B(\xi) = \sum_j \xi(x_j)B(x_j)$  donde  $B(x_j) = 1/c_j$ . Y el parámetro  $\beta$  se ajustará en cada ejemplo.

La expresión (3.6) del algoritmo multiplicativo queda en este caso de L-optimización:

$$\xi^{(s+1)}(x_j) = \xi^{(s)}(x_j) \frac{\beta B(x_j) - \text{tr} \{ -M^{-1}(\xi^{(s)}) L M^{-1}(\xi^{(s)}) M(1_{x_j}) \}}{\beta \sum_j \xi^{(s)}(x_j) B(x_j) - \text{tr} \{ -M^{-1}(\xi^{(s)}) L \}}.$$

De (3.2) la FIM es

$$M(\xi) = H^T \left( \Sigma^{-1} \otimes \sum_j \xi_j x_j^T x_j \right) H$$

donde  $\xi_j = \xi(x_j)$ , de modo que  $\sum_j \xi_j = 1$ . De (3.4)  $L = H^T (I \otimes \sum_j x_j^T x_j) H$ , que es una matriz definida positiva de la misma dimensión que  $M(\xi)$ .

Tomando el parámetro  $\beta$  del intervalo  $[0, 500]$  para el primer ejemplo se obtiene  $\beta^* = 66,875$  con un tamaño del diseño aproximado de  $n^{(\beta^*)} = 332,08$ . El diseño exacto (redondeado) óptimo se presenta en la Tabla 3.1, denotado con P. El valor de la función objetivo para el diseño aproximado

es  $\Phi_L [M(n^{(\beta^*)}\xi^{(\beta^*)})] = 0,021223$ , mientras que para el diseño exacto es 0,021222, ligeramente mejor.

Para el segundo ejemplo se considera  $\beta \in [0, 2000]$  y se obtiene  $\beta^* = 586,5$ . El diseño exacto se presenta en la Tabla 3.3 y alcanza 446 experimentos, un 16 % menos que el diseño exacto de Conlisk, que propone 529, pero con un coste similar. Además,  $\Phi_L [M(n^{(\beta^*)}\xi^{(\beta^*)})] = 0,0907064$  para el diseño aproximado, siendo 0,091391 para el exacto.

Las Tablas 3.1 y 3.3 muestran grandes diferencias entre los diseños obtenidos por los dos planteamientos. En los diseños obtenidos con nuestro procedimiento hay numerosos ceros, 4 de 9 en el ejemplo simplificado, concretamente en los puntos pares de soporte, y 26 de 36 en el ejemplo completo. En ambas versiones los diseños se concentran en los puntos de soporte de los ‘extremos’ y el centro, como es habitual en los modelos de regresión. Como comportamiento general, ambos planteamientos asignan el mayor número de experimentos en el punto de soporte central, el más barato.

### Ventajas del procedimiento

Hay dos ventajas destacables del procedimiento aquí propuesto. Por un lado el ahorro de tiempo de computación. El problema (3.4) y nuestro procedimiento se han implementado en *Mathematica*<sup>®</sup>. Para obtener cualquiera de los Casos 1, 2 ó 3 se han necesitado en torno a 60 segundos, mientras que nuestro procedimiento apenas ha consumido 6 segundos. Obviamente este asunto es crucial cuando el número de puntos de soporte del diseño crece. Baste señalar que en el ejemplo de Conlisk (1979) se consideran 36 condiciones experimentales (puntos) y nuestro procedimiento necesita sólo unos 34 segundos para obtener el diseño óptimo.

Por otro lado, los diseños obtenidos con nuestro procedimiento tienen grandes eficiencias. Para comparar los diseños se calculan las L-eficiencias relativas (véase la Sección 1.9)

$$\text{eff}_L(\xi, \xi') = \frac{\Phi_L[M(\xi')]}{\Phi_L[M(\xi)]},$$

donde se consideran los diseños con su tamaño total de experimentos  $n$  ó  $n^{(\beta)}$  lo que implica que las magnitudes de las matrices de información varían dependiendo de dicho tamaño total. La L-eficiencia relativa del diseño aproximado con 9 puntos de soporte obtenido con nuestro procedimiento respecto a los casos 1, 2 ó 3 es del 98,8 %. Si se consideran los diseños exactos la L-eficiencia cae al 98,6 % en el peor de los casos. Por otro lado, el diseño exacto de 36 puntos obtenido con nuestro procedimiento es un 26,6 % más eficiente que el diseño estructural proporcionado por Conlisk. Nótese que ambos diseños son L-óptimos restringidos, obtenidos a igualdad de costes y no para tamaños de muestra iguales. De hecho, los diseños obtenidos serían los aplica-

bles en la práctica por lo que en el cálculo de la L-eficiencia no debe dividirse por el tamaño de muestra.

Las dos ventajas destacadas se basan en dos puntos. Por un lado, el uso de un algoritmo multiplicativo eficiente. Por otro lado, la manejable definición de la matriz asintótica de covarianzas de los estimadores de los parámetros estructurales del modelo. Además, el procedimiento establecido es aplicable a cualquier problema de diseño con restricciones.

### El problema del redondeo

Otro asunto que se ha señalado como importante es el uso de diseños aproximados cuando la solución requiere diseños exactos. Los diseños aproximados obtenidos, tanto resolviendo el problema de programación no lineal (3.4) como con nuestro procedimiento, deben redondearse y un método de redondeo, de los muchos que hay (Pukelsheim, 2014), puede conducir a un diseño exacto mejor que otro. Como complemento al procedimiento descrito, los diseños obtenidos se han redondeado utilizando los 9 métodos de redondeo que proporciona el software BAZI (Maier y Pukelsheim, 2007). Por ejemplo, para el diseño estructural de 36 puntos de soporte el método de cuota de Hare proporciona un reparto de experimentos con menor  $\Phi_L(\xi)$  que el redondeo estándar, también implementado en BAZI. La diferencia entre ambos se encuentra en dos puntos de soporte. En  $P_g = P_r = D_g = D_r = 1$ , donde el método de cuota de Hare asigna un experimento menos que el redondeo estándar, de 166 a 165, que se ‘mueve’ a  $P_g = P_r = 1,07; D_g = D_r = 0$ , donde el método de Hare asigna 27 experimentos por los 26 del redondeo estándar. Por lo tanto, el último paso del procedimiento será escoger el método que mejor redondeo proporciona basándose en el criterio estadístico considerado.

Los nueve métodos de prorrateo de BAZI se dividen en siete métodos de división y dos métodos de cuota. Balinski y Young (2001) señalan que los métodos de división son muy superiores a los de cuota.

Todos los métodos de división constan de dos etapas: escalado y redondeo. En la etapa de escalado los números/votos iniciales se escalan por un divisor común. La etapa de redondeo depende del método de división, y consiste en ajustar el divisor para completar el tamaño total del prorrateo, número de escaños, prefijado de antemano. Los siete métodos de división corresponden a dos familias. Por un lado la familia potencia-media caracterizada por el parámetro  $0 \neq p \in \mathbb{R}$  y las señales

$$s_p(n) = \left( \frac{(n-1)^p + n^p}{2} \right)^{1/p}$$

que dividen cada intervalo  $[n-1, n]$  en dos partes. Los números, votos o proporciones iniciales se dividen por el divisor y cuando el cociente resultante cae en el intervalo  $[n-1, n]$ , se redondea a la baja, a  $n-1$ , si cae por debajo de la señal. Cuando el cociente cae por encima de la señal se redondea

al alza, a  $n$ . BAZI implementa esta familia de métodos, permitiendo elegir cualquier  $p \in \mathbb{R}$ . Ahora bien, los cinco métodos de división ‘tradicionales’ según Balinski y Young (2001) son casos particulares de esta familia y BAZI los implementa directamente. La diferencia entre los cinco métodos es el redondeo: a la baja, cuyos contribuidores originales fueron Jefferson, D’Hondt y Hagenbach–Bischoff y que equivale a considerar  $p \rightarrow +\infty$ ; al alza debido a Adams que se obtiene tomando  $p \rightarrow -\infty$ ; redondeo estándar, debido a Webster, Sainte–Laguë y Schepers, cuando  $p = 1$ ; redondeo armónico, propuesto por Dean, tomando  $p = -1$  y redondeo geométrico también llamado método de proporciones iguales, de Hill y Huntington, que corresponde al caso límite  $p \rightarrow 0$ .

Por otro lado están los métodos de división de la familia con redondeo estacionario. En este caso el parámetro es  $q \in [0, 1]$  y la señal  $s_q(n) = n - 1 + q$ . Si  $q = 0, 1/2, 1$  se tienen respectivamente los métodos de división con redondeo al alza, estándar o a la baja. Para los cálculos de estos métodos de división BAZI utiliza un algoritmo de baja complejidad computacional (véase Happacher y Pukelsheim, 1996; Dorffleitner y Klein, 1999).

Los dos métodos de cuota incluidos en BAZI son el método de cuota de Hare (Hamilton/Hare/Niemeyer) y el método de cuota de Droop, ambos con ajuste residual mediante restos mayores. La cuota de Hare es el cociente, sin redondear, entre el número total y el tamaño total del reparto. La cuota de Droop es igual a 1 más la parte entera del número total dividido por 1 más el tamaño total del reparto. De aquí que el método de Droop siempre sea un entero positivo.

## Capítulo 4

# Conclusiones y líneas de trabajo futuro

En este capítulo se enumeran las principales conclusiones de esta tesis, estructuradas en torno a los tres problemas considerados. Como fruto ya constatable de este trabajo se encuentran dos publicaciones y un artículo enviado para su publicación:

- Casero–Alonso V. y López–Fidalgo J. (2014). Experimental designs in triangular simultaneous equations. *Statistical Papers*.  
DOI: 10.1007/s00362-014-0581-y
- Stehlík M., López–Fidalgo J., Casero–Alonso V. y Bukina E. (2014). Robust integral compounding criteria for trend and correlation structures. *Stochastic Environmental Research and Risk Assessment*.  
DOI: 10.1007/s00477-014-0892-5
- Casero–Alonso V. y López–Fidalgo J. (2014). Optimal designs subject to cost constraints in simultaneous equations models. *submitted*.

que se adjunta en el capítulo de anexos. Además estos resultados se han presentado en diversos congresos tanto nacionales como internacionales. Concretamente, en los tres congresos nacionales de Jóvenes Investigadores en Diseño de Experimentos y Bioestadística habidos hasta la fecha: JEDE<sup>1</sup> en 2010 en Toledo, JEDE II<sup>2</sup> en 2012 en Tenerife y JEDE III<sup>3</sup> en 2014 en Pamplona. En dos congresos nacionales de la Sociedad de Estadística e Investigación Operativa: SEIO 2012<sup>4</sup> en Madrid y SEIO 2013<sup>5</sup> en Castellón. Y en la XIII

---

<sup>1</sup><http://areaestadistica.uclm.es/jede/>

<sup>2</sup><https://sites.google.com/site/jedetenerife/home>

<sup>3</sup><http://www.unavarra.es/actualidad/congresos?contentId=179923>.

<sup>4</sup><http://www.seio2012.com/>

<sup>5</sup><http://www.seio2013.com/>

y XIV Conferencia Española de Biometría celebradas en 2011 en Barcelona y en 2013 en Ciudad Real<sup>6</sup>.

En cuanto a los congresos internacionales se han presentado ponencias en SPO 11 (The Pyrenees International Workshop and Summer School on Statistics, Probability and Operations Research) en 2011 en Jaca, en Statistische Woche 2012<sup>7</sup> en Vienna (Austria) y en ENBIS-14<sup>8</sup> (14th Annual Conference of the European Network for Business and Industrial Statistics) en 2014 en Linz (Austria).

A los congresos internacionales DAE 2012<sup>9</sup> (USA), ERCIM 2012<sup>10</sup> (España), y mODa 10<sup>11</sup> (Polonia) se han presentado trabajos por invitación, recibiendo además 'A DAE 2012 Student Scholarship Award'. En DAE 2012 se expuso un póster y en ERCIM 2012 y mODa 10 se presentaron ponencias. Del congreso mODa: Model-Oriented Data Analysis and Optimum Design, ya se ha dicho que reúne a investigadores en el campo del diseño óptimo de experimentos, mayoritariamente de Europa. La participación es por rigurosa invitación combinando investigadores seniors con nóveles en un número limitado en torno a 80 participantes y sin sesiones paralelas. Similarmente el congreso DAE: Design and Analysis of Experiments, reúne mayoritariamente a investigadores seniors y nóveles de USA. El congreso ERCIM sin embargo acoge numerosas presentaciones, en torno a 800 en la edición de 2012, organizadas en sesiones ordinarias e invitadas, combinando estadística y computación en áreas como métodos robustos, bioestadística, métodos bayesianos, tratamiento de datos longitudinales y funcionales, series temporales, diseño óptimo, estadística computacional, etc.

### Conclusiones del problema 1: Diseños experimentales en modelos triangulares de ecuaciones simultáneas

1. *La modelización mediante ecuaciones simultáneas, SE, del problema de búsqueda del diseño óptimo con variables no controlables, es un planteamiento más general que la modelización vía diseños condicionalmente restringidos, CR.* La diferencia entre ambos planteamientos estriba en suponer para la variable no controlable, una distribución de probabilidad parcialmente conocida a través de unos parámetros estimables, enfoque SE, o un diseño condicional conocido, enfoque CR. No obstante, si la distribución de probabilidad en el planteamiento SE fuese totalmente conocida de antemano, no hubiese que estimar nada en ella, ambos planteamientos conducirían a los mismos diseños. Si los

---

<sup>6</sup><http://areaestadistica.uclm.es/biometria2013/>

<sup>7</sup><http://www.statistische-woche.at>

<sup>8</sup>[http://www.enbis.org/activities/events/current/253\\_ENBIS\\_14/index?\\_ts=146951](http://www.enbis.org/activities/events/current/253_ENBIS_14/index?_ts=146951)

<sup>9</sup><http://www.stat.uga.edu/events/DAE>

<sup>10</sup><http://www.cfe-csda.org/ercim12/>

<sup>11</sup><http://www.moda10.uz.zgora.pl/>

parámetros de una de las ecuaciones simultáneas no tienen nada en común con los de la otra ecuación, la matriz de información del modelo SE será diagonal por cajas (Teorema 3.1.) Además el planteamiento SE permite capturar la interacción entre las variables endógenas.

2. *El diseño CR D-óptimo para el modelo considerado cuando  $\chi = \{0, 1\}$  depende de los parámetros  $r_0$  y  $r_1$  de la distribución condicional conocida  $\tilde{\xi}_{x|z}$  y puede ser de dos tipos (véase la Figura 3.1 izquierda). En la región a la derecha de la curva  $r_0 = 2r_1(1 - r_1)$  el soporte del diseño está formado por los dos puntos del espacio de diseño,  $x = 0$  y  $x = 1$ . En la otra región, el diseño es unipuntual, con soporte sólo en el punto  $x = 1$ . Además los pesos son simétricos en el parámetro  $r_1$  respecto del valor  $1/2$ . Esto es, se tiene el mismo diseño para  $(r_0^0, r_1^0)$  y  $(r_0^0, 1 - r_1^0)$ .*
3. *Existe un sólo tipo de diseño SE D-óptimo para el modelo considerado cuando  $\chi = \{0, 1\}$ , que también depende de los parámetros  $r_0$  y  $r_1$  que determinan la distribución de la variable no controlable. El soporte del diseño óptimo está formado por los dos puntos del espacio de diseño. Para realizar los cálculos se ha modelizado  $z$  mediante la distribución de Bernoulli equiparando los parámetros estimables a los valores de  $r_0$  y  $r_1$ . Como se deduce del Teorema 3.1, los diseños a los que se llega por la vía SE son distintos, se obtienen distintos pesos, a los del enfoque CR por suponer que los parámetros son desconocidos y estimables. Los pesos también son simétricos en  $r_1$  y están acotados entre 0,5 y 0,75.*
4. *Las eficiencias de los diseños SE respecto a los diseños CR siempre son mayores que a la inversa ( $\chi = \{0, 1\}$ ). Hay que tener en cuenta que en la obtención de los primeros han considerado 2 parámetros más que en la obtención de los segundos. No obstante, dependiendo de los valores de  $r_0$  y  $r_1$  las eficiencias de los diseños obtenidos por una vía respecto de la otra pueden ser muy parecidas. O bien totalmente distintas, como ocurre en la región sombreada de la Figura 3.1 (derecha) donde los diseños CR son ineficientes desde el punto de vista del planteamiento SE.*
5. *El diseño SE D-óptimo para el modelo considerado es bastante robusto ( $\chi = \{0, 1\}$ ). Los diseños obtenidos con unos parámetros nominales cualesquiera son muy eficientes aunque los verdaderos parámetros se alejen considerablemente de ellos. Las eficiencias quedan por encima del 90 % en la mayoría de los casos. De hecho, ante el desconocimiento de los valores reales de los parámetros, la mejor opción es tomar como valores nominales el punto medio  $r_0 = r_1 = 1/2$ , que minimiza la distancia a cualquier punto de la región definida para ellos, para el que la eficiencia relativa respecto al diseño con cualesquiera parámetros reales no baja del 97 %. Al ser simétricos en  $r_1$ , un diseño con*

parámetros nominales  $(r_0^0, r_1^0)$  será 100 % eficiente para dos conjuntos de parámetros reales con valores de  $(r_0^0, r_1^0)$  y  $(r_0^0, 1 - r_1^0)$ .

6. *Existen tres regiones de diseños CR D-óptimos para el modelo lineal considerado con espacio de diseño  $\chi = [0, 1]$*  (véase la Figura 3.2 izquierda). Salvo una pequeña región, los diseños obtenidos en el caso  $\chi = \{0, 1\}$  son también D-óptimos para el espacio de diseño continuo. El nuevo tipo de diseño, de la región superior izquierda, se concentra en el punto  $0 < t^* < 1$ . Los tres diseños obtenidos están por debajo del límite de puntos de soporte que proporciona el corolario del Teorema de Caratheodory, 3 en este caso. En esta extensión se ha supuesto una modelización de  $z$  mediante una distribución Bernoulli de parámetro  $r_x$  lineal. Para otros supuestos de  $r_x$  los diseños D-óptimos en el espacio discreto no son óptimos cuando se extiende el espacio al intervalo continuo.

7. *Existen diversas regiones de diseños SE D-óptimos para  $\mathcal{X} = [0, 1]$  dependiendo de la distribución asumida para la variable no controlable.* Suponiendo una función lineal  $r_x$  para modelizar la probabilidad de  $z = 1$ , se encuentran cuatro regiones de diseños SE D-óptimos de tres tipos distintos de diseños, con soporte en un punto interior  $u$  o en dos, bien los dos extremos  $\{0, 1\}$  o bien dos puntos interiores  $\{u, t\}$  (Figura 3.2 derecha). Los diseños SE D-óptimos con el espacio de diseño discreto sólo son D-óptimos en un entorno de la diagonal principal.

Suponiendo funciones no lineales los resultados son muy distintos. Concretamente, suponiendo una distribución logística, conforme nos alejamos de la diagonal principal donde los diseños SE D-óptimos son del tipo  $\{0, T\}$ , los diseños óptimos evolucionan desde diseños de dos puntos del tipo  $\{t^*, 1\}$  a diseños con soporte en tres puntos del tipo  $\{t^*, u^*, T\}$  (véase la Figura 3.3).

8. *Es más arriesgado tomar valores nominales en unas regiones que en otras ( $\chi = [0, 1]$ ).* Ante el desconocimiento de los parámetros reales del modelo y suponiendo  $r_x$  lineal, en el enfoque CR ha de evitarse la región superior izquierda, mientras que en el SE hay que evitar el punto medio. La región a evitar cambia con  $r_x$ , por ejemplo, tomando una distribución logística se han de evitar las regiones superior izquierda e inferior derecha.

9. *Pseudo-paradoja diseño exacto-aproximado.* Se ha obtenido un diseño D-óptimo aproximado con un único punto de soporte del diseño. El correspondiente diseño D-óptimo exacto siempre tiene dos puntos de soporte, independientemente del tamaño del diseño. Siempre diferirán en su aplicación práctica.

### Conclusiones del problema 2: Criterio robusto de composición integral

1. *Se introduce una nueva clase de criterios de composición integral.* Se consiguen con ellos diseños robustos respecto de dos o más enfoques para un problema dado, por ejemplo modelos rivales o partes diferentes de un modelo (tendencia y correlación, o ecuaciones de un modelo de ecuaciones simultáneas). Se componen en forma integral las respectivas matrices de información, FIMs, que se optimizan respecto a una función de densidad.
2. *Los criterios compuestos clásicos pueden no ser adecuados para obtener diseños robustos.* Ponen al mismo nivel las dos FIMs lo que puede ser inapropiado.
3. *Mediante un kernel se introduce la discrepancia entre las dos FIMs.* Permitirá poner a distinto nivel las dos FIMs. Algunos kernels producen el mismo espectro de diseños óptimos (Lema 1, Sección 1.2 anexo A.2).
4. *El criterio compuesto clásico así como los criterios para discriminar entre modelos rivales son casos particulares del criterio de composición integral definido.* Si en el criterio de composición integral la densidad utilizada para optimizar no depende de los parámetros del modelo(s), se tiene el criterio compuesto clásico (Lema 2 Sección 1.2 anexo A.2). Criterios como T- o KL-optimización, teniendo en cuenta algunas generalizaciones, se pueden incluir en el criterio de composición integral definido.
5. *Se obtiene un diseño robusto en el caso del modelo de ecuaciones simultáneas aplicando el criterio de composición integral.* Se considera un kernel basado en eficiencias y la distribución Beta para optimizar el criterio. Se obtiene un diseño con una eficiencia en torno al 87% respecto de las dos partes (ecuaciones) del modelo SE.

### Conclusiones del problema 3: Diseños con restricciones de costes

1. *Se define un procedimiento para obtener diseños óptimos con restricciones de costes para modelos de ecuaciones simultáneas.* Se basa en un algoritmo multiplicativo aplicado a un criterio compuesto convexo. El criterio combina mediante un parámetro escalar una función criterio estadística con una función de costes. Para una mejor aplicación del algoritmo se ha considerado una función de beneficios apropiada basada en la de costes. Además el GET permite calcular la eficiencia de los diseños en cada paso del algoritmo pudiendo escoger el grado de afinidad al óptimo.

2. *El procedimiento descrito puede utilizarse para resolver problemas de diseño con restricciones para cualquier criterio estadístico convexo.* Existe dependencia de los parámetros al tratarse de un modelo no lineal. Para el ejemplo de 9 puntos de soporte no se ha encontrado un único mínimo global solución del problema de programación no lineal.
3. *El procedimiento descrito reduce el tiempo de cálculo respecto a la resolución como problema de programación no lineal.* La reducción, en torno a 10 veces más rápido en el ejemplo considerado, radica en el empleo del algoritmo multiplicativo y en la manejable expresión de la matriz de información derivada de Rothenberg y Leenders (1964). En el problema 1 no se consideró dicha expresión para la matriz de información.
4. *Los diseños obtenidos con el procedimiento son muy eficientes.* Para el ejemplo de L-optimización considerado, el diseño obtenido con el procedimiento es un 26,6% más L-eficiente que el proporcionado por Conlisk (1979).
5. *Los pesos de diseño de ambos planteamientos son bastante diferentes.* En los obtenidos con nuestro procedimiento hay numerosos ceros concentrándose el diseño en los puntos de soporte de los ‘extremos’ y el centro, como es habitual en los modelos de regresión. Ambos planteamientos asignan el mayor número de experimentos en el punto de soporte central, el más barato.
6. *El redondeo juega un papel fundamental.* En el ejemplo considerado, como en muchos casos reales, los diseños deben ser exactos. El procedimiento descrito utiliza diseños aproximados que deben redondearse. Un método de redondeo puede ser mejor que otro en el sentido de proporcionar valores mejores de la función objetivo.

## Líneas de trabajo futuro

Algunas líneas futuras de investigación son las siguientes:

1. Plantear problemas de diseño en Modelos dosis-respuesta (p.ej. Biswas y López-Fidalgo, 2013) desde la perspectiva de los diseños para modelos de ecuaciones simultáneas.
2. Considerar en el criterio de composición integral el caso en el que los parámetros de los modelos rivales o de las diferentes partes consideradas del modelo sean comunes.
3. Generalizar el criterio de composición integral para espacios de diseños arbitrarios de  $n$  puntos de soporte considerando una densidad de Dirichlet para los pesos de diseño.

4. Ampliar el problema de diseño con restricciones analizando la sensibilidad respecto a la elección de los valores nominales de los parámetros tanto estructurales como de correlación.
5. Desarrollar algoritmos para diseñar experimentos con modelos estructurales con muchas variables endógenas y exógenas.
6. Plantear el problema del diseño óptimo en modelos de ecuaciones estructurales en el campo médico con variables confusoras, modeladoras y mediadoras, en colaboración con la Unidad de Trastornos Alimentarios del complejo hospitalario de Ciudad Real.



# Bibliografía

- [1] Abdelbasit K.M. y Plackett R.L. (1983). Experimental design for binary data. *Journal of the American Statistical Association*, **78**, 381, 90–98.
- [2] Ahmadi J. y Arghami N.R. (2003). Comparing the Fisher information in record values and iid observations. *Statistics: A Journal of Theoretical and Applied Statistics*, **37(5)**, 435–441.
- [3] Aigner D.J. (1979). A brief introduction to the methodology of optimal experimental design. *Journal of Econometrics*, **11(1)**, 7–26.
- [4] Aigner D.J. y Balestra P. (1988). Optimal experimental design for error components models. *Econometrica*, **56(4)**, 955–971.
- [5] Alshunnar F.S., Raqab M.Z. y Kundu D. (2012). On the Comparison of the Fisher Information of the Log–Normal and Generalized Rayleigh Distributions. *Journal of Applied Statistics*, **37(3)**, 391–404.
- [6] Amato U. y Hughes W. (1991). Maximum entropy regularization of Fredholm integral equations of the first kind. *Inverse Problems* **7**, 793–808.
- [7] Amo–Salas M., López–Fidalgo J. y Porcu E. (2013). Optimal designs for some stochastic processes whose covariance is a function of the mean. *Test*, **22**, 159–181.
- [8] Angrist J.D. y Pischke J.S. (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press, Princeton, New Jersey.
- [9] Ardanuy R. y López–Fidalgo J. (1992). Optimal design with constraint support. *Revista de matematica e statistica*, **10**, 193–206.
- [10] Atkinson A.C. y Donev A.N. (1992). *Optimum experimental designs*. Oxford Science Publications. Oxford.
- [11] Atkinson A.C., Donev A.N. y Tobias R.D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press. Oxford.

- [12] Atkinson A.C. y Fedorov V.V. (1975a). The designs of experiments for discriminating between two rival models. *Biometrika*, **62**, 57–70.
- [13] Atkinson A.C. y Fedorov V.V. (1975b). Optimal design: experiments for discriminating between several models. *Biometrika*, **62**, 289–303.
- [14] Atkinson A.C. y Hunter W.G. (1968). The design of experiments for parameter estimation. *Technometrics*, **10(2)**, 271–289.
- [15] Atwood C.L. (1976). Convergent design sequences, for sufficiently regular optimality criteria. *The Annals of Statistics*, **4(6)**, 1124–1138.
- [16] Baldi Antognini A. y Zagoraiou M. (2010). Exact optimal designs for computer experiments via Kriging metamodeling. *Journal of Statistical Planning and Inference*, **140**, 2607–2617
- [17] Balinski M. y Young P. (2001). *Fair Representation. Meeting the Ideal of One Man, One Vote. Second Edition*. Brookings Institution Press: Washington, DC.
- [18] Barzilai J. y Borwein J.M. (1988). Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, **8**, 141–148.
- [19] BAZI. Berechnung von Anzahlen mit Zuteilungsmethoden im Internet / Calculation of Allocations by Apportionment Methods in the Internet. <http://www.uni-augsburg.de/bazi/>
- [20] Bentler P.M. y Stein J.A. (1992). Structural equation models in medical research. *Statistical Methods in Medical Research*, **1(2)**, 159–181.
- [21] Berger M.P.F. y Wong W.K., editores (2005). *Applied Optimal Designs*. Wiley.
- [22] Berger M.P.F. y Wong W.K. (2009). *An Introduction to Optimal Designs for Social and Biomedical Research*. John Wiley & Sons.
- [23] Biswas A. y López-Fidalgo J. (2013). Compound designs for dose-finding in the presence of nondesignable covariates. *Pharmaceutical Statistics* **12(2)**, 92–101.
- [24] Box G.E.P. (1979). Some problems of statistics and everyday life. *Journal of the American Statistical Association*, **74**, 1–4.
- [25] Box G.E.P. y Hunter W.G. (1965). Sequential design of experiments for non-linear models. *Proceedings of the IBM Scientific computing symposium in statistics*, 113–137.
- [26] Box G.E.P. y Lucas H.L. (1959). Design of experiments in non-linear situations. *Biometrika*, **46**, 77–90.

- 
- [27] Box M.J. (1968a). The use of designed experiments in nonlinear model-building. *The Future of Statistics. Watts D.G. (Ed.), Academic Press, Inc., New York*, **241**, 257.
- [28] Box M.J. (1968b). The occurrence of replications in optimal designs of experiments to estimate parameters in nonlinear models. *Journal of the Royal Statistical Society, Series B*, **30**, 290–302.
- [29] Box M.J. (1969). *The design of experiments and nonlinear model-building under a variety of error structures*. Ph.D. Thesis. University of London, London, England.
- [30] Box M.J. (1970). Some experiences with a nonlinear experimental design criterion. *Technometrics*, **12(3)**, 569–589.
- [31] Brown L.D., Olkin I., Sacks J. y Wynn H.P., editores (1985). *Jack Carl Kiefer Collected Papers III. Design of Experiments*. Springer-Verlag. New York.
- [32] Casero-Alonso V. y López-Fidalgo J. (2014). Experimental designs in triangular simultaneous equations models. *Statistical Papers (in press)*. DOI: 10.1007/s00362-014-0581-y
- [33] Casero-Alonso V. y López-Fidalgo J. (2014). Optimal designs subject to cost constraints in simultaneous equations models. *Submitted*
- [34] Coetzer R.L.J y Haines L.M. (2013). Optimal Designs for Multiple-Mixture by Process Variable Experiments. *In: Uciński D., Atkinson A.C. y Patan M., editores (2013). mODa 10: Advances in Model-Oriented Design and Analysis*. Contributions to Statistics. Springer International Publishing Switzerland.
- [35] Conlisk J. (1979). Design for simultaneous equations. *Journal of Econometrics*, **11(1)**, 63–76.
- [36] Cook R.D. y Fedorov V.V. (1995). Constrained optimization of experimental design. *Statistics: A Journal of Theoretical and Applied Statistics*, **26(2)**, 129–178.
- [37] Cook R.D. y Wong W.K. (1994). On the Equivalence of Constrained and Compound Optimal Designs. *Journal of the American Statistical Association*, **89(426)**, 687–692
- [38] Cornell J. A. (2002). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data (3rd ed.)*. Wiley. New York.
- [39] Crary S.B. (2002). Design of Computer Experiments for Metamodel Generation. *Analog Integrated Circuits and Signal Processing*, **32**, 7–16.

- [40] Demidenko E. (2004). *Mixed models. Theory and applications*. Wiley.
- [41] Dette H., Pepelyshev A. y Zhigljavsky A. (2014). ‘Nearly’ universally optimal designs for models with correlated observations. *Computational Statistics & Data Analysis*, **71(C)**, 1103-1112.
- [42] Dorfleitner G. y Klein T. (1999). Rounding with multiplier methods: An efficient algorithm and applications in statistics. *Statistical Papers*, **20**, 143–157.
- [43] Dragalin V. y Fedorov V. (2006). Design of multi-centre trials with binary response. *Statistics in Medicine*, **25**, 2701–2719
- [44] Draper N.R. y Hunter W.G. (1967). The use of prior distributions in the design of experiments for parameter estimation in non-linear situations: multiresponse case. *Biometrika*, **54**, 662–665.
- [45] Draper N.R. y Pukelsheim F. (1996). An overview of design of experiments. *Statistical Papers*, **37(1)**, 1–32
- [46] Ehrenfeld S. (1955). On the efficiency of experimental designs. *The Annals of Mathematical Statistics*, **26**, 247–255.
- [47] Elfving G. (1952). Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics*, **23**, 255–262.
- [48] Fedorov V.V. (1972). *Theory of optimal experiments*. Department of statistics Purdue University.
- [49] Fedorov V.V. y Hackl P. (1997). *Model-Oriented Design of Experiments*. Lecture Notes in Statistics. Springer, New York.
- [50] Fedorov V.V. y Leonov S.L. (2013). *Optimal Design for Nonlinear Response Models*. Chapman & Hall/CRC Biostatistics Series
- [51] Ford I., Titterton D.M. y Kitsos Ch.P. (1989). Recent Advances in Nonlinear Experimental Design. *Technometrics*, **31(1)**, 49–60.
- [52] García J.L., Rivera L., Blanco J., Jiménez E. y Martínez E. (2014). Structural equations modelling for relational analysis of JIT performance in maquiladora sector. *International Journal of Production Research*, (ahead-of-print)  
DOI:10.1080/00207543.2014.885143
- [53] Goos P. y Jones B. (2011). *Optimal design of experiments: a case study approach*. John Wiley & Sons.
- [54] Gujarati D. (1997). *Econometría básica*. McGraw–Hill. ISBN 958–600–585–2

- [55] Hahn J., Keisuke H. y Dean K. (2011). Adaptive Experimental Design using the Propensity Score. *Journal of Business & Economic Statistics*, **29**(1), 96–108.
- [56] Hansen P.C. (1992). Numerical tools for analysis and solution of Fredholm integral equations of the first kind. *Inverse Problems*, **8**, 849–872.
- [57] Happacher M. y Pukelsheim F. (1996). Rounding probabilities: Unbiased multipliers. *Statistics & Decision*, **14**, 373–382.
- [58] Harman R. (2008). Equivalence theorem for Schur optimality of experimental designs. *Journal of Statistical Planning and Inference*, **138**, 1201–1209.
- [59] Hill D.H. (1980). D-optimal designs for partially nonlinear regression models. *Technometrics*, **22**(2), 275–276.
- [60] Hoel P.G. (1958). Efficiency problems in polynomial estimation. *The Annals of Mathematical Statistics*, **29**, 1134–1145.
- [61] Hofmann G. (2004). Comparing the Fisher information in record data and random observations. *Statistical Papers*, **45**(4), 517–528.
- [62] Hu F. y Rosenberger W.F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley & Sons
- [63] Imhof L., López-Fidalgo J. y Wong W.K. (2001). Efficiencies of rounded optimal approximate designs for small samples. *Statistica Neerlandica*, **55**, 301–315.
- [64] Karlin S. y Studden W.I. (1966). Optimal experimental designs. *The Annals of Mathematical Statistics*, **37**, 783–810.
- [65] Kessels R., Jones B., Goos P. y Vandebroek M. (2009). An efficient algorithm for constructing Bayesian optimal choice designs. *Journal of Business & Economic Statistics*, **27**(2), 279–291.
- [66] Khuri A.I. y Lee J. (1998). A graphical approach for evaluating and comparing designs for nonlinear models. *Computational Statistics & Data Analysis*, **27**(4), 433–443.
- [67] Kiefer J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society, Series B*, **21**, 272–319.
- [68] Kiefer J. (1974). General equivalence theory for optimum designs (Approximate theory). *The Annals of Statistics*, **2**, 5, 848–879.
- [69] Kiefer J. y Wolfowitz J. (1959). Optimum design in regression problems. *The Annals of Mathematical Statistics*, **30**, 271–294.

- [70] Kiefer J. y Wolfowitz J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, **12**, 363–366.
- [71] Kirkegaard H., Stovring H., Rasmussen K.M., Abrams B., Sørensen T.I. y Nohr E.A. (2014). How do pregnancy-related weight changes and breastfeeding relate to maternal weight and BMI-adjusted waist circumference 7 y after delivery? Results from a path analysis. *The American Journal of Clinical Nutrition*, **99**(2), 312–319
- [72] Kiselák J. y Stehlík M. (2008). Equidistant and D-optimal designs for parameters of Ornstein–Uhlenbeck process. *Statistics & Probability Letters*, **78**(12), 1388–1396.
- [73] Kitsos Ch.P. (2013). *Optimal Experimental Design for Non-Linear Models*. Springer Briefs in Statistics. Springer.
- [74] Levy S. y Steinberg D.M. (2010). Computer experiments: a review *Advances in Statistical Analysis*, **94**(4), 311–324
- [75] Lin C.D., Anderson-Cook C.M., Hamada M.S., Moore L.M. y Sitter R.R. (2014). Using Genetic Algorithms to Design Experiments: A Review. Quality and Reliability Engineering International. DOI: 10.1002/qre.1591
- [76] López-Fidalgo J. (2009). A critical overview on optimal experimental designs. *Boletín de Estadística e Investigación Operativa*, **25**(1), 14–21.
- [77] López-Fidalgo J. (2011). *Diseño óptimo de experimentos*. UCLM <http://www.uclm.es/profesorado/jesuslopezfidalgo/optimal.pdf>
- [78] López-Fidalgo J. y Garcet-Rodríguez S. (2004). Optimal experimental designs when some independent variables are not subject to control. *Journal of the American Statistical Association*, **99**, 1190–1199.
- [79] López-Fidalgo J. y Rivas-López M.J. (2007). MV-Optimality Standardized through the Coefficient of Variation. *Journal of Statistical Planning and Inference*, **137**, 2680–2007.
- [80] López-Fidalgo J., Rivas-López M.J. y Fernández-Garzón, B. (2007). A-Optimality Standardized Through the Coefficient of Variation. *Communications in Statistics. Theory and Methods*, **36**(4), 781–792.
- [81] López-Fidalgo J., Rodríguez-Díaz J.M., y Torsney B., editores (2007). *mODa8: Advances in Model-Oriented Design and Analysis*. Contributions to Statistics. Springer International Publishing Switzerland.
- [82] López-Fidalgo J., Tommasi Ch. y Trandafir C. (2007). An optimal experimental design criterion for discriminating between non-Normal

- models. *Journal of the Royal Statistical Society, Series B*, **69**(2), 231–242.
- [83] López-Fidalgo J., Tommasi Ch. y Trandafir C. (2008). Optimal designs for discriminating between some extensions of the Michaelis–Menten model. *Journal of Statistical Planning and Inference*, **138**, 3797–3804.
- [84] Madden G. (1993). Australian telephone network subscription and calling demands: evidence from a stated–preference experiment. *Information Economics and Policy*, **5**, 207–230.
- [85] Madden G. (1995). Experimentation in economics: An overview of the stated–preference experimental design method. *Australian Economic Papers*, **34**(64), 120–135.
- [86] Maier S. y Pukelsheim F. (2007). Bazi – A free computer program for proportional representation apportionment. *Preprint Nr. 42/2007. Institut für Mathematik, Universität Augsburg*.  
<http://opus.bibliothek.uni-augsburg.de/volltexte/2007/711/>
- [87] Mandal A., Wong, W. K., y Yu, Y. (2014). Algorithmic Searches for Optimal Designs. *Handbook of Design and Analysis of Experiments*. Chapman & Hall/CRC.
- [88] Mandal B., Roe B.E. y Fein S.B. (2014). Work and breastfeeding decisions are jointly determined for higher socioeconomic status US mothers. *Review of Economics of the Household*, **12**(2), 237–257.
- [89] Martín-Martín R., Torsney B. y López-Fidalgo J. (2007). Construction of marginally and conditionally restricted designs using multiplicative algorithms. *Computational Statistics & Data Analysis*, **51**, 5547–5561.
- [90] Masten M.(2014). Random coefficients on endogenous variables in simultaneous equations models. *Cemmap working paper, Centre for Microdata Methods and Practice*, No. CWP01/14.
- [91] McGree J.M., Eccleston J.A. y Duffull S.B. (1988). Compound optimal design criteria for nonlinear models. *Journal Of Biopharmaceutical Statistics*, **18**(4), 646–661.
- [92] Melas V.B. (2005). *Functional Approach to Optimal Experimental Design*. Lecture Notes in Statistics. Springer.
- [93] Moler J.A., Plo F. and San Miguel M. (2006). An adaptive design for clinical trials with non–dichotomous response and prognostic factors. *Statistics & Probability Letters*, **76**, 1940–1946.
- [94] Montgomery D.C. (2001). *Design and analysis of experiments*. John Wiley & Sons, Inc.

- [95] Müller W.G. y Pronzato L. (2009). Towards an optimal design equivalence theorem for random fields? *IFAS report Nr. 45 of the Department for Applied Statistics of the Johannes Kepler University in Linz*.
- [96] Müller W.G. y Stehlík M. (2009). Issues in the Optimal Design of Computer Simulation Experiments. *Applied Stochastic Models in Business and Industry*, **25**, 163–177.
- [97] Müller W.G. y Stehlík M. (2010). Compound optimal spatial designs. *Environmetrics*, **21**, 354–364.
- [98] Newey W.K., Powell J.L. y Vella F. (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, **67**(3), 565–603.
- [99] NGuyen T.T., Bazzoli C. y Mentré F. (2012). Design evaluation and optimisation in crossover pharmacokinetic studies analysed by nonlinear mixed effects models. *Statistics in Medicine*, **31** (11–12), 1043–1058
- [100] Novales A. (2002). *Econometría. 2ª ed.* McGraw–Hill. Madrid.
- [101] Papakyriazis P.A. (1986). Adaptive optimal estimation control strategies for systems of simultaneous equations. *Mathematical Modelling*, **7**, 241–257.
- [102] Pázman A. (1986). *Foundations of optimum experimental design*. D. Reidel publishing company. Dordrecht.
- [103] Pázman A. (2010). Information Contained In Design Points Of Experiments With Correlated Observations. *Kybernetika*, **46**(4), 771–783
- [104] Pistone G., Riccomagno E. y Wynn H.P. (2000). *Algebraic Statistics: Computational Commutative Algebra in Statistics*. Chapman & Hall.
- [105] Porra R.J. (2002). The chequered history of the development and use of simultaneous equations for the accurate determination of chlorophylls a and b. *Photosynthesis research*, **73**, 149–156.
- [106] Poskitt D.S. y Skeels C.L. (2008). Conceptual frameworks and experimental design in simultaneous equations. *Economics Letters*, **100**, 138–142.
- [107] Pronzato L. (2010). Penalized optimal designs for dose–finding. *Journal of Statistical Planning and Inference*, **140**, 283–296.
- [108] Pronzato L. y Pázman A. (2013). *Design of Experiments in Nonlinear Models*. Lecture Notes in Statistics. Springer.

- [109] Pukelsheim F. (1993). *Optimal Design of Experiments*. John Wiley & Sons. New York.
- [110] Pukelsheim F. (2014). *Proportional Representation: Apportionment Methods and Their Applications*. Springer International Publishing Switzerland.
- [111] Pukelsheim F. y Rieder S. (1992). Efficient rounding of approximate designs. *Biometrika*, **79**(4), 763–770.
- [112] Ranjan P., Bingham D. y Michailidis G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, **50**(4), 527–541.
- [113] Rasch D., Pilz J., Verdooren L.R. y Gebhardt A. (2011). *Optimal Experimental Design with R*. Chapman and Hall/CRC.
- [114] Regúlez M. (2008). *Ecuaciones Simultáneas con aplicaciones en Gretl*. UPV.
- [115] Rodríguez-Díaz J.M., Santos-Martín M.T., Waldl H. y Stehlik M. (2012). Filling and D-optimal designs for the correlated generalized exponential models. *Chemometrics and Intelligent Laboratory Systems*, **114**, 10–18
- [116] Rodríguez Torreblanca C. y Ortíz Rodríguez I. (1999). *Diseño Óptimo de Experimentos para Modelos de Regresión*. Universidad de Almería, Servicio de Publicaciones, Almería.
- [117] Rothenberg T.J. y Leenders C.T. (1964). Efficient estimation of simultaneous equations systems. *Econometrica*, **32**, 57–76.
- [118] Sacks J., Schiller S.B. y Welch W.J. (1989). Design for Computer Experiment. *Technometrics*, **31**(1), 41–47.
- [119] Schwabe R. (1996). *Optimum Designs for Multi-Factor Models*. Springer, New York.
- [120] Silvey S.D. (1980). *Optimal design*. Chapman & Hall.
- [121] Silvey S.D. y Titterington D.M. (1973). A geometric approach to optimal design theory. *Biometrika*, **60**(1), 21–32.
- [122] Silvey S.D., Titterington D.M. y Torsney B. (1978). An algorithm for optimal designs on a finite space. *Communications in Statistics - Theory and Methods* **14**, 1379–1389.

- [123] Smith K. (1918). On the standard deviations of adjusted and interpolates values of an observed polynomial functions and its constraints and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, **12**, 1–85.
- [124] Stehlík M., López-Fidalgo J., Casero-Alonso V. y Bukina E. (2014). Robust integral compounding criteria for trend and correlation structures. *Stochastic Environmental Research and Risk Assessment*. DOI: 10.1007/s00477-014-0892-5
- [125] Tandeo P., Ailliot P. y Autret E. (2011). Linear Gaussian state-space model with irregular sampling: application to sea surface temperature. *Stochastic Environmental Research and Risk Assessment*, **25**, 793–804
- [126] Tikhonov A.N. y Arsenin V.Y. (1977). *Solutions of Ill-posed Problems*. Wiley, New York.
- [127] Titterton D.M. (1976). Algorithms for computing D-optimal designs on a finite design space. *Proceedings of the 1976 Conference on Information Sciences and Systems, Johns Hopkins University*, 213–216
- [128] Torsney B. y Martín-Martín R. (2009). Multiplicative algorithms for computing optimum designs. *Journal of Statistical Planning and Inference*, **139**, 3947–3961.
- [129] Uciński D. (2005). *Optimal Measurement Methods for Distributed Parameter System Identification*. CRC Press LLC.
- [130] Uciński D., Atkinson A.C. y Patan M., editores (2013). *mODa 10: Advances in Model-Oriented Design and Analysis*. Contributions to Statistics. Springer International Publishing Switzerland.
- [131] Unami K., Abagale F.K., Yangyuoru M., Alam A.H.M.B. y Kranjac-Berisavljevic G. (2010). A stochastic differential equation model for assessing drought and flood risks. *Stochastic Environmental Research and Risk Assessment*, **24**, 725–733.
- [132] Varela G., Cordovilla R., Jiménez M.F. y Novoa N. (2001). Utility of Standardized Exercise Oximetry to Predict Cardiopulmonary Morbidity After Lung Resection. *European Journal of Cardiothoracic Surgery*, **19**, 351–354.
- [133] Wahba G. (1977). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM Journal on Numerical Analysis*, **14**, 651–67
- [134] Wald A. (1943). On the efficient design of statistical investigations. *The Annals of Mathematical Statistics*, **14**, 134–140.

- 
- [135] Welch W.J. (1982). Branch-and-bound search for experimental designs based on D-optimality and other criteria. *Technometrics*, **24(1)**, 41–48
- [136] Whittle P. (1973). Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society, Series B*, **1**, 123–130.
- [137] White L.V. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika*, **60(2)**, 345–348.
- [138] Wilmut M. y Zhou J. (2011). D-optimal minimax design criterion for two-level fractional factorial designs. *Journal of Statistical Planning and Inference*, **141**, 576–587.
- [139] Woods D.C. (2010). Robust designs for binary data: applications of simulated annealing. *Journal of Statistical Computation and Simulation*, **80**, 29–41.
- [140] Wu Ch.-F. y Wynn H.P. (1978). The convergence of general step-length algorithms for regular optimum design criteria. *The Annals of Statistics*, **6(6)**, 1273–1285.
- [141] Wynn H.P. (1970). The sequential generation of D-optimum experimental designs. *The Annals of Mathematical Statistics*, **41(5)**, 1655–1664
- [142] Zellner A. y Theil H. (1962). Three-Stage least squares: simultaneous estimation of simultaneous equations. *Econometrica*, **30**, 54–78.
- [143] Zhigljavsky A.A., Dette H. y Pepelyshev A. (2010). A new approach to optimal design for linear models with correlated observations. *Journal of the American Statistical Association*, **105**, 1093–1103.
- [144] Zhigljavsky A.A., Pronzato L. y Bukina E. (2013). An asymptotically optimal gradient algorithm for quadratic optimization with low computational cost. *Optimization Letters*, **7(6)**, 1047–1059.



# Anexos



# Experimental designs in triangular simultaneous equations models

Víctor Casero-Alonso · Jesús López-Fidalgo

Received: 12 April 2013 / Revised: 20 January 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** Optimal experimental designs are considered for models with simultaneous equations. In particular, a model with two equations is assumed where one of the explanatory variables (exogenous) of the first equation is then the response variable (endogenous) of the second equation. In both equations there is a control variable, which is being designed through the celebrated D-optimality criterion. This work is based on a more restricted approach using just the first equation and assuming the distribution of the exogenous/endogenous variable completely known. Then a conditionally restricted optimal design was computed afterwards. In this paper the conditional model is assumed partially known but it has to be fitted as well. Although both approaches identify different prior knowledge a comparison of the optimal designs for both approaches is made. Since the model is not linear in the usual sense the optimal designs will depend on the parameters and a sensitivity analysis against its choice is performed.

**Keywords** Approximate design · Information matrix · D-optimal design · Simultaneous equations · Conditionally restricted design · Structural equations

**Mathematics Subject Classification** 62K05, 62P20

## 1 Introduction

The experimental design theory can be applied in a wide variety of contexts. As [Draper and Pukelsheim \(1996\)](#) pointed out, this topic could well have an entire encyclopaedia

---

V. Casero-Alonso (✉) · J. López-Fidalgo  
Department of Mathematics, Institute of Mathematics Applied to Science and Engineering,  
University of Castilla-La Mancha, Avda. Camilo José Cela 3, 13005 Ciudad Real, Spain  
e-mail: VictorManuel.Casero@uclm.es

devoted to it. On the other hand, the list of examples of simultaneous equations (SE) problems is quite extensive in economics and other areas. The literature about experimental designs in economics is not extensive, although increasing numbers of experiments are being conducted in economics. A special issue of the *Journal of Econometrics* acknowledged the significance of experimental design in econometrics. For instance, in this issue [Aigner \(1979\)](#) discussed the optimal allocation of observations for multivariate regression and [Conlisk \(1979\)](#) considered oil pricing experiments using SE models. He introduced an objective function based either on the three stage least squares or on the full information maximum likelihood estimators. Though the optimal design must be exact, his procedure is to approximate it by solving the continuous programming problem and rounding off. Other examples are provided by [Aigner and Balestra \(1988\)](#) in electricity pricing experiments, [Kessels et al. \(2009\)](#) in conjoint choice experiments or [Hahn et al. \(2011\)](#) in social experiments.

In this paper we consider a linear regression model with explanatory variables some of which are not subject to the control of the practitioner. Moreover we consider that the uncontrolled variables have unknown values before the experiment is carried out. [López-Fidalgo and Garcet-Rodríguez \(2004\)](#) considered this case for the duration of an exercise test to obtain more information to predict complications of surgery in the treatment of lung cancer taking into account the oxygen desaturation during the test. They obtained optimal experimental designs considering a unique equation, with prior information about the variable not subject to control with unknown values before the experiment is performed. This approach is called conditionally restricted (CR) optimality. As a different approach a SE model is considered in this paper. The variable not under control is considered as explanatory first and as a response in a second equation, forming a triangular model.

It is worth mentioning that [Angrist and Pischke \(2009\)](#) have frequently encouraged the use of a single endogenous regressor. This is in line with not entertaining overly complex models. Another example of triangular SE is a dose–response model where toxicity is exogenous in the equation of the efficacy while it is endogenous in a second equation, where the efficacy is not present ([Biswas and López-Fidalgo 2013](#)). This is because it is expected that toxicity will affect efficacy, but not the opposite. An example of triangular SE in economics where an experimental design may be optimized is [Newey \(1999\)](#). They investigated the empirical non-linear relationship between the hourly wage rate and the number of annual hours worked.

The organization of this paper is as follows. In Sect. 2 the general background and the two different approaches of the problem are presented. In Sect. 3 D-optimal designs for a two-point design space are obtained. In Sect. 4 the design space is extended to the continuous case. In both sections the two approaches are considered separately and D-optimal designs are compared by calculating relative efficiencies. We conclude with a discussion in Sect. 5.

## 2 Approach to the problem

### 2.1 Optimal experimental design theory

A statistical model relates the responses,  $y$ , and explanatory variables,  $x$ , through a parametric family of probability density functions (pdf),  $f(y|x, \theta)$ . The aim of this theory is to find experimental conditions  $x \in \chi$ , where  $\chi$  is a compact design space, so that the parameters can be efficiently estimated. An approximate design is defined as a probability measure on  $\chi$ , say  $\xi$  (e.g. [Kiefer 1974](#)). The Fisher Information Matrix (FIM) of a design,  $\xi$ , may be defined as

$$M(\xi) = \int_{\chi} E \left[ -\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta^T} \right] \xi(dx),$$

where  $L(\theta) = \log f(y|x, \theta)$  is the log-likelihood function for a particular experiment.

For uncorrelated observations the covariance matrix of the maximum likelihood estimator (MLE) of  $\theta$  is asymptotically  $\sigma^2 n^{-1} M^{-1}(\xi)$ . Thus, efficient estimation of this model is closely connected to the properties of the information matrices. Using Caratheodory's Theorem we can restrict the search to finite and bounded support designs,

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_k \\ p_1 & p_2 & \dots & p_k \end{array} \right\},$$

where  $x_1, \dots, x_k$  are distinct points in  $\chi$  forming the design support, and  $p_i$  are the associated design weights ( $0 \leq p_i \leq 1, \sum_i p_i = 1$ ). In practice, if  $n$  experiments are going to be realized approximately  $np_i$  of them will be made at  $x_i$ .

There is a wide variety of optimality criteria which may be appropriate for designing an experiment. A criterion function is a convex real function  $\Phi[M(\xi)]$  to be minimized. In this work  $\Phi_D[M(\xi)] = |M(\xi)|^{-1/m}$  is used as a criterion function, where  $m$  is the number of parameters of the model. This criterion, called D-optimality is the most popular and minimizes the volume of the confidence ellipsoid of the parameters.

The general equivalence theorem (GET) ([Kiefer 1974](#)) provides conditions to verify the optimality of a design and so it provides a lower bound of the efficiency of a design

$$\gamma = \text{eff}_{\Phi}(\xi) = \frac{\Phi[M(\xi^*)]}{\Phi[M(\xi)]},$$

where  $\xi^*$  is the  $\Phi$ -optimal design.

More details about this theory may be found, e.g. at [Pazman \(1986\)](#) and [Atkinson et al. \(2007\)](#).

### 2.2 CR and SE theory

The objective of this paper is to find optimal designs for estimating a model with two explanatory variables,  $x$  and  $z$ . It is assumed that we can design only the variable  $x$ ,

while  $z$  is not subject to control. However, the values of  $z$  are known only after the experiment is performed.

Let  $x = (x_1, \dots, x_n)^T$  be  $n$  design points and  $y = (y_1, \dots, y_n)^T$ ,  $z = (z_1, \dots, z_n)^T$  random vectors such that the conditional distribution of  $y|z$  is given by a pdf  $g(y|z; x, \theta)$  with mean  $\eta_{12}(z; x, \theta)$  and  $\text{var}(y|z) = \sigma_{12}^2(z; x, \theta)$ .

López-Fidalgo and Garcet-Rodríguez (2004) considered a linear regression model where the response variable depends on two explanatory variables. One of them is controlled, but the other is uncontrolled and unknown before the experiment is realized. They proposed a one-equation model  $E(y|z) = \eta_{12}(z; x, \theta)$ , assuming there is prior information about the unknown values of  $z$  given  $x$  through a known conditional design (probability distribution)  $\tilde{\xi}_{z|x}(\cdot|x)$ . The optimal design in this case is called CR (López-Fidalgo and Garcet-Rodríguez 2004; Martín-Martín et al. 2007).

We propose alternatively a model of two SE. Thus, we introduce the second equation to model  $z$ , in terms of  $x$ , by means of the marginal distribution of  $z$  given by a pdf  $h(z; x, \theta)$ . So the SE model may be expressed as

$$\begin{aligned} y &= \eta_{12}(z; x, \theta) + \varepsilon_{12} \\ z &= \eta_2(x, \theta) + \varepsilon_2, \end{aligned}$$

where  $E(\varepsilon_{12}|z) = 0$ ,  $\text{var}(\varepsilon_{12}|z) = \text{var}(y|z) = \sigma_{12}^2(z; x, \theta)$ ,  $E(\varepsilon_2) = 0$  and  $\text{var}(\varepsilon_2) = \text{var}(z) = \sigma_2^2(x, \theta)$ . The case of different parameters in both equations is included here, e.g.  $\theta = (\theta_1, \theta_2)^T$ ,  $\theta_1$  for the first equation and  $\theta_2$  for the second. This second model can capture interactions between variables (Conlisk 1979). Thus it is considered a more general approach.

Another way of representing this is

$$\begin{aligned} y &= \eta_1(x, \theta) + \varepsilon_1 \\ z &= \eta_2(x, \theta) + \varepsilon_2, \end{aligned}$$

where  $E(\varepsilon_1) = 0$ ,  $\text{var}(\varepsilon_1) = \text{var}(y) = E[\sigma_{12}^2(z; x, \theta)] + \text{var}[\eta_{12}(z; x, \theta)]$ ,  $E(y) = \eta_1(x, \theta) = E[\eta_{12}(z; x, \theta)]$  and  $\text{cov}(\varepsilon_1, \varepsilon_2) = \text{cov}(y, z) = E[z\eta_{12}(z; x, \theta)] - \eta_2(x, \theta)E[\eta_{12}(z; x, \theta)]$ . This may be called the reduced form (RF). The triangular structure in the SE model allows an easy transformation to the RF. The later may be redefined with transformed parameters, but the original ones are those of interest. Conlisk (1979) uses the RF with the transformed parameters and then computes the L-optimal design<sup>1</sup> using for that the transformation matrix of the parameters. We use the RF with the original parameters. Then we obtain the FIM deriving with respect to them and finally we compute the D-optimal design. Although D-optimality is invariant for this transformation the problem has essentially the same difficulty. For this reason we chose the original parameters to preserve those of interest throughout the process. In the example considered in the next section the original parameters are identifiable. In other cases identifiability has to be checked carefully.

<sup>1</sup> The minimization of the criterion function  $\Phi[M(\xi)] = \text{tr}\{M^{-1}(\xi)L\}$ , with  $L$  a  $p \times q$  matrix of coefficients, leads to a L-optimal design (Atkinson et al. 2007).

The log-likelihood function is

$$L(\theta) = \log [g(y|z; x, \theta)h(z; x, \theta)],$$

and maximizing it the MLE are obtained.

The information matrix for a joint design,  $\xi(x, z)$ , is

$$M(\theta, \xi) = \sum_{x,z} \xi(x, z)I(\theta; x, z) = \sum_{x,z} \xi(x, z)E_{y|x,z} \left[ -\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta^T} \right], \tag{1}$$

where  $I(\theta; x, z)$  is the FIM at point  $(x, z)$ .

A different, but obvious, RF, say RF2, would be the result of plugging the second equation into the first one,

$$\begin{aligned} y &= \eta_{12}(\eta_2(x, \theta); x, \theta) + \varepsilon_3 \\ z &= \eta_2(x, \theta) + \varepsilon_2, \end{aligned}$$

where  $E(\varepsilon_3) = 0$ ,  $\text{var}(\varepsilon_3) = \text{var}(y) = E[\sigma_{12}^2(z; x, \theta)] + \text{var}[\eta_{12}(z; x, \theta)]$  and  $\text{cov}(y, z) = E[z\eta_{12}(z; x, \theta)] - \eta_2(x, \theta)E[\eta_{12}(z; x, \theta)]$ . Then a new joint distribution must be considered. But this is a different model except in the case of  $\eta_{12}$  linear on  $z$  (an example is provided in the Sect. 5).

From this the following theorems hold.

**Theorem 1** *With the previous general conditions,*

$$\begin{aligned} E(y) &= E[\eta_{12}(z; x, \theta)], \\ \text{var}(y) &= E[\text{var}(y|z)] + \text{var}[\eta_{12}(z; x, \theta)], \\ \text{cov}(y, z) &= E[z\eta_{12}(z; x, \theta)] - \eta_2(x, \theta)E[\eta_{12}(z; x, \theta)]. \end{aligned}$$

*If  $\eta_{12}(z; x, \theta) = \alpha(x, \theta) + \beta(x, \theta)z$  and  $\text{var}(y|z) = \tau^2$ , then*

$$\begin{aligned} E(y) &= \eta_{12}(\eta_2(x, \theta); x, \theta) = \alpha(x, \theta) + \beta(x, \theta)\eta_2(x, \theta), \\ \text{var}(y) &= \tau^2 + \beta(x, \theta)^2\text{var}(z), \\ \text{cov}(y, z) &= \beta(x, \theta)\text{var}(z). \end{aligned}$$

*Proof* It is straightforward handling the expectations. □

**Theorem 2** *If  $M_{CR}$  is the information matrix of a CR design  $\xi_{CR}$ ,  $M_{SE}$  is the information matrix of the corresponding SE design  $\xi_{SE}$  and  $\theta_1$  and  $\theta_2$  are the parameters of the first and second equation, respectively, then:*

1. *If  $h(z; x, \theta_2) = \tilde{\xi}_{z|x}(\cdot|x)$  and is fully known beforehand:  $M_{SE} = M_{CR}$  and then  $\xi_{SE}^* = \xi_{CR}^*$  for any optimality criteria.*

2. If  $\theta_1$  and  $\theta_2$  are unknown under the SE approach and they have no components in common, then

$$M_{SE}(\theta, \xi_{SE}) = \begin{pmatrix} M_{CR}(\theta_1, \xi_{SE}) & 0 \\ 0 & M_{II}(\theta_2, \xi_{SE}) \end{pmatrix},$$

where  $M_{II}$  is the information matrix of the second equation. This diagonal block structure is a consequence of the triangular structure of the SE model.

For simplicity of notation we use  $\xi_{CR}$  ( $\xi_{SE}$ ) both for the joint design  $\xi_{CR}(x, z)$  ( $\xi_{SE}(x, z)$ ) and for the marginal design  $\xi_{CR}(x)$  ( $\xi_{SE}(x)$ ), whenever there is no possible confusion.

*Proof* First of all, we will obtain the information matrix for the two approaches. From (1) it follows that, for a CR design:

$$M_{CR}(\theta_1, \xi_{CR}) = \sum_{x,z} \xi_{CR}(x) \tilde{\xi}_{z|x}(z|x) E_{y|z;x} \left[ -\frac{\partial^2 \log g(y|z; x, \theta_1)}{\partial \theta_1 \partial \theta_1^T} \right]$$

because  $\xi_{CR}(x, z) = \xi_{CR}(x) \tilde{\xi}_{z|x}(z|x)$ .

And for a SE design:

$$M_{SE}(\theta, \xi_{SE}) = \sum_x \xi_{SE}(x) E_z \left[ E_{y|z} \left( -\frac{\partial^2 [\log g(y|z; x, \theta_1) + \log h(z; x, \theta_2)]}{\partial \theta \partial \theta^T} \right) \right]$$

because the joint distribution of  $y$  and  $z$  is  $f(y, z; x, \theta) = g(y|z; x, \theta_1) h(z; x, \theta_2)$ .

1. If  $h(z; x, \theta_2) = \tilde{\xi}_{z|x}(\cdot|x)$  and it is fully known beforehand, then  $\theta_2$  does not need to be estimated and

$$\frac{\partial^2 L(\theta)}{\partial \theta_1 \partial \theta_1^T} = \frac{\partial^2 \log g(y|z; x, \theta_1)}{\partial \theta_1 \partial \theta_1^T}.$$

thus  $M_{SE} = M_{CR}$ .

2. If  $\theta_1$  and  $\theta_2$  have no components in common:

$$L(\theta) = \log g(y|z; x, \theta_1) + \log h(z; x, \theta_2) = L_1(\theta_1) + L_2(\theta_2),$$

from which the result is obtained.

□

### 3 D-Optimal designs for a two-point design space

We propose two paradigmatic examples, which can be viewed as representatives of the general problem. In the medical example given in the introduction, the standard

times for the exercise are 12 and 18 min. In the oil pricing example we can experiment with two prices. Then, the two-point design space may be considered as a first step of the analysis. Later on we will consider the continuous design space case.

### 3.1 CR D-optimal design

López-Fidalgo and Garcet-Rodríguez (2004) used the following linear regression model to illustrate the CR D-optimal design method:

$$E(y|z; x) = \eta_{12}(z; x, \theta_1) = \alpha x + \beta z,$$

where the response variable depends on  $x \in \chi = \{0, 1\}$  for two possible modalities of the experiment and  $z \in \mathcal{Z} = \{0, 1\}$  with two possible outcomes of the experiment. Any prior information about  $z$  is given by:

$$\tilde{\xi}_{z|x}(\cdot|0) = \begin{Bmatrix} 0 & 1 \\ 1 - r_0 & r_0 \end{Bmatrix}, \quad \tilde{\xi}_{z|x}(\cdot|1) = \begin{Bmatrix} 0 & 1 \\ 1 - r_1 & r_1 \end{Bmatrix}.$$

This means that it is known (e.g., from retrospective data) that about  $100r_0\%$  of the outcomes are one when we know that  $x$  is zero and about  $100r_1\%$  of the outcomes are one when we know that  $x$  is one. Thus,  $z$  has a Bernoulli distribution for each value of  $x$  with known mean  $r_x$  ( $x = 0, 1; r_x \in [0, 1]$  excluding meaningless cases:  $(r_0, r_1) = (0, 0), (0, 1), (1, 0), (1, 1)$ ).

The CR optimal design will be a joint design of  $x$  and  $z$  subject to the conditional distribution  $\tilde{\xi}_{z|x}$ :

$$\xi_{CR}(x, z) = \xi_{CR}(x)\tilde{\xi}_{z|x}(z|x) = \begin{Bmatrix} (0, 0) & (0, 1) & (1, 0) & (1, 1) \\ (1 - p)(1 - r_0) & (1 - p)r_0 & p(1 - r_1) & pr_1 \end{Bmatrix},$$

where we will optimize:

$$\xi_{CR}(x) = \begin{Bmatrix} 0 & 1 \\ 1 - p & p \end{Bmatrix}, \quad p \in [0, 1].$$

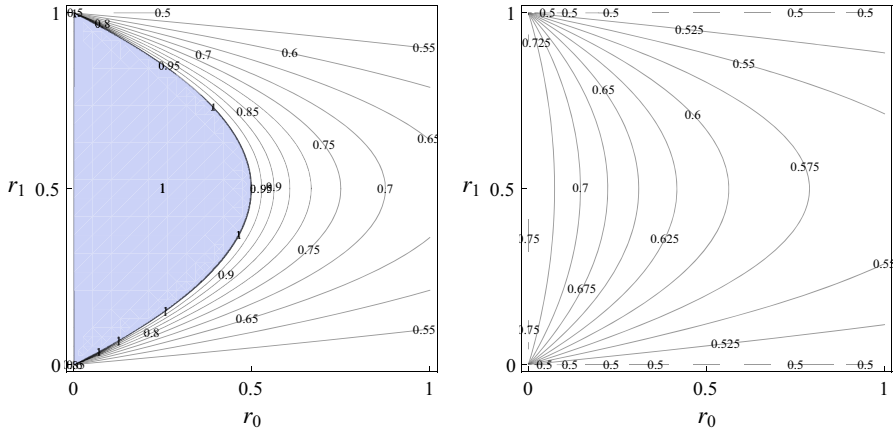
That is, since the design space has only two support points, we choose the weight of the support point  $x = 1: \xi_{CR}(1) = p$  to optimize according to a particular criterion.

The information matrix of the design is:

$$M_{CR}(\theta_1, \xi_{CR}) = \begin{pmatrix} p & pr_1 \\ pr_1 & (1 - p)r_0 + pr_1 \end{pmatrix}.$$

Thus, the CR D-optimal weights (and therefore the CR D-optimal designs), which maximize the determinant of the information matrix, are:

$$p^* = \xi_{CR}^*(1) = \begin{cases} \frac{r_0}{2[r_0 - r_1(1 - r_1)]} & r_0 > 2r_1(1 - r_1), \\ 1 & r_0 \leq 2r_1(1 - r_1). \end{cases}$$



**Fig. 1** Values of  $p^*$  for CR D-optimal designs (left) and SE D-optimal designs (right) for  $\chi = \{0, 1\}$  and  $(r_0, r_1) \in [0, 1] \times [0, 1]$

The CR D-optimal designs depend on the nominal values of the parameters and are symmetric with respect to  $r_1$ . Figure 1 (left) shows the values of  $p^*$  for CR D-optimal designs based on the values of  $r_0$  and  $r_1$ .

In the paper of López-Fidalgo and Garcet-Rodríguez (2004) there is a numerical example to illustrate the CR approach:  $\xi_{z|x}(1|0) = r_0 = 3/4$  and  $\xi_{z|x}(1|1) = r_1 = 7/8$  [a point in the right hand region of Fig. 1 (left)] then  $\xi_{CR}^*(1) = p^* = 24/41 = 0.585$ . That is, if it is known that about 75 % of the outcomes are one when  $x$  is zero and 87.5 % of the outcomes are one when  $x$  is one then 58.5 % of the experiments must be for  $x = 1$  to achieve minimization of the volume of the confidence ellipsoid of the parameters  $\alpha$  and  $\beta$  (CR D-optimal design for  $x$ ).

We compute two other numerical examples for later comparison. One of them is  $r_0 = 2/3$  and  $r_1 = 3/4$ , again from the right hand side of Fig. 1 (left). In this case the D-optimal weight is  $p^* = 16/23 = 0.696$ . The third numerical example, taken from the left hand side of Fig. 1 (left) is  $r_0 = 1/3$  and  $r_1 = 1/2$  and then  $p^* = 1$ .

### 3.2 SE D-optimal design

As mentioned above, we consider a SE model with two equations. We choose a model corresponding to the example used for illustrating the CR approach:

$$\begin{cases} E(y|z; x) = \eta_{12}(z; x, \theta_1) = \alpha x + \beta z, \\ E(z|x) = \eta_2(x; \theta_2) = r_x. \end{cases}$$

The unknown variable  $z$  is modeled by the probability of  $z = 1$ . However, since we assume  $\chi = \{0, 1\}$ , only  $r_0$  and  $r_1$  are to be determined. We suppose  $\mathcal{Z} = \{0, 1\}$  and the random variable  $y$  has a normal distribution. Hence, the joint distribution is:

$$\begin{aligned}
 f(y, z|x; \theta) &= g(y|z; x, \theta_1) h(z; x, \theta_2) \\
 &= \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\alpha x)^2} (1 - r_x) & \text{if } z = 0, \\ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\alpha x-\beta)^2} r_x & \text{if } z = 1. \end{cases}
 \end{aligned}$$

Since we assume that the parameters  $\alpha$  and  $\beta$  are different from the parameters  $r_0$  and  $r_1$  of Theorem 2 it follows that the information matrix of the design is:

$$M_{SE}(\theta, \xi_{SE}) = \left( \begin{array}{c|cc} M_{CR}(\theta_1, \xi_{SE}) & & 0 \\ \hline 0 & \frac{-1+p}{(-1+r_0)r_0} & 0 \\ & 0 & \frac{p}{r_1-r_1^2} \end{array} \right). \tag{2}$$

After some algebra, the SE D-optimal design obtained is:

$$\xi_{SE}^*(x) = \left\{ \begin{array}{cc} 0 & 1 \\ 1 - p_{SE}^* & p_{SE}^* \end{array} \right\}.$$

where  $\xi_{SE}^*(1) = p_{SE}^* = \arg \max_p |M_{SE}| = \frac{6r_0 + 3(r_1 - 1)r_1 - \sqrt{D}}{8(r_0 + (r_1 - 1)r_1)}$ , and  $D = 4r_0^2 + 4r_0(r_1 - 1)r_1 + 9(r_1 - 1)^2 r_1^2$ . In this case  $p_{SE}^* \in [0.5, .75]$  for all  $(r_0, r_1) \in [0, 1] \times [0, 1]$  [see Fig. 1 (right)]. As for the CR D-optimal design, the SE D-optimal design depends on the nominal values and is symmetric with respect to  $r_1$ . But now, we only have one kind of SE D-optimal designs for all possible values of  $r_0$  and  $r_1$  while we had two types of CR D-optimal designs.

### 3.3 Comparisons, relative efficiencies and robustness

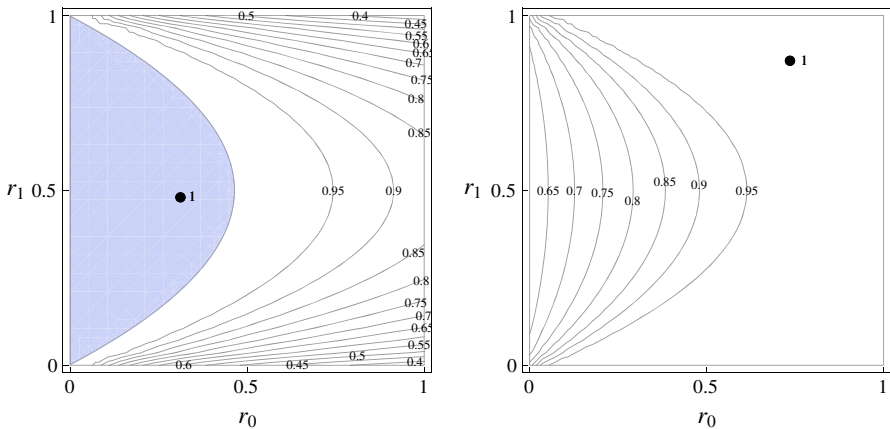
The CR and SE D-optimal designs for the numerical examples considered above are shown in Table 1. For the same parameter values the CR and SE D-optimal design are different. But the real comparisons are in terms of relative efficiencies: the efficiency of the CR D-optimal design relative to the SE model and the efficiency of SE D-optimal design relative to the CR model are

$$\text{eff}_{SE}(\xi_{CR}^*) = \left( \frac{|M_{SE}(\theta, \xi_{CR}^*)|}{|M_{SE}(\theta, \xi_{SE}^*)|} \right)^{1/4} \quad \text{and} \quad \text{eff}_{CR}(\xi_{SE}^*) = \left( \frac{|M_{CR}(\theta_1, \xi_{SE}^*)|}{|M_{CR}(\theta_1, \xi_{CR}^*)|} \right)^{1/2}.$$

The efficiencies for the numerical examples considered above are presented in Table 1. The first conclusion from these results is that CR D-optimal design for the numerical example 3 is inefficient for the SE approach. This conclusion is valid for all the designs of the region on the left hand side of the curve  $r_0 = 2r_1(1 - r_1)$ . The second conclusion is that efficiencies of the SE D-optimal designs for the CR model are greater than efficiencies of CR D-optimal designs for the SE model. This is a logical result because the SE model has four parameters and the CR model has two.

**Table 1** CR and SE D-optimal designs and relative efficiencies in percentages

$(r_0, r_1)$	$\xi_{CR}^*(1)$	$\xi_{SE}^*(1)$	$eff_{SE}(\xi_{CR}^*)$	$eff_{CR}(\xi_{SE}^*)$
$(3/4, 7/8)$	.585	.536	99.57	99.64
$(2/3, 3/4)$	.696	.567	97.21	98.28
$(1/3, 1/2)$	1	.644	0	84.87



**Fig. 2** Efficiencies of the CR D-optimal design with nominal parameters  $(r_0^0, r_1^0) = (1/3, 1/2)$  (left) and  $(r_0^0, r_1^0) = (3/4, 7/8)$  (right) for all possible true values of  $r_0$  and  $r_1$

Nevertheless in some cases the efficiencies are quite similar, as in the first and the second numerical example.

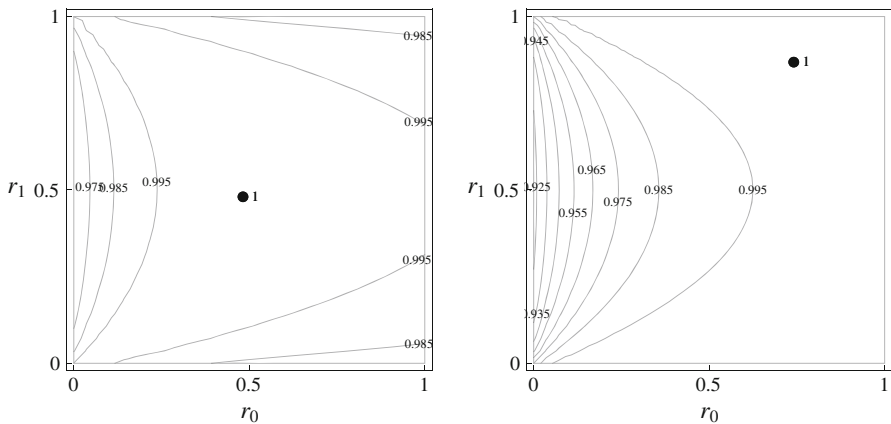
Since the CR and SE designs depend on the nominal values of the parameters a robustness analysis will be performed. This will be measured by the efficiency

$$eff_{\theta^*}(\xi_{\theta^0}^*) = \left( \frac{|M(\theta^*, \xi_{\theta^0}^*)|}{|M(\theta^*, \xi_{\theta^*}^*)|} \right)^{1/m}, \tag{3}$$

where  $\theta^0$  are the nominal parameters and  $\theta^*$  different possible true values of the parameters. This efficiency indicates the goodness of the CR or SE D-optimal design assuming nominal parameters  $r_0^0$  and  $r_1^0$  when the true parameters are  $r_0^*$  and  $r_1^*$ .

Various scenarios of nominal and true parameters can be considered. For the CR D-optimal design if  $r_0$  and  $r_1$  are nominal values satisfying  $r_0 \leq 2r_1(1 - r_1)$ , that is,  $(r_0, r_1)$  is a point on the left hand side of Fig. 1 (left), the efficiency is 100 % for all the designs with true parameters in the same region. For the remaining possible true values (right hand side) the efficiency decreases dramatically to values below 50 % [see Fig. 2 (left)].

For nominal values on the right hand side of Fig. 1 (left) we have the opposite behaviour as Fig. 2 (right) shows. However we have a smaller drop in efficiency; in the worst case it is just above 60 %, for nominal values  $(r_0^0, r_1^0) = (3/4, 7/8)$ . As a matter of fact, the efficiency of the CR D-optimal design obtained with these nominal values, assuming  $(1/3, 1/2)$  as true values, is 81.6 % whereas the efficiency when the nominal values are  $(1/3, 1/2)$  assuming  $(3/4, 7/8)$  as true values is 70.6 %.



**Fig. 3** Efficiencies of the SE D-optimal design with nominal parameters  $(r_0^0, r_1^0) = (1/2, 1/2)$  (left) and  $(r_0^0, r_1^0) = (3/4, 7/8)$  (right) for all possible true values of  $r_0$  and  $r_1$

For the SE D-optimal design, the most conservative scenario is  $(r_0^0, r_1^0) = (1/2, 1/2)$  because these are the values which minimize the distance to any point  $(r_0, r_1) \in [0, 1] \times [0, 1]$ . The efficiencies are shown in Fig. 3 (left). The first conclusion from this figure is that, in the worse case, the efficiency is greater than 97 %. So, this conservative scenario is the best when there is no information about the true parameters. Thus the SE D-optimal design is quite robust. For comparisons, Fig. 3 (right) shows the efficiencies of the SE D-optimal design for nominal values  $(r_0^0, r_1^0) = (3/4, 7/8)$ . Notice that, in this case, the efficiency is greater than 92 %.

Another conclusion is that these efficiencies are symmetric with respect to  $r_1$ , that is, the CR and SE D-optimal designs have the same efficiency for true parameters  $r_1^*$  or  $1 - r_1^*$  for fixed  $r_0^*$ .

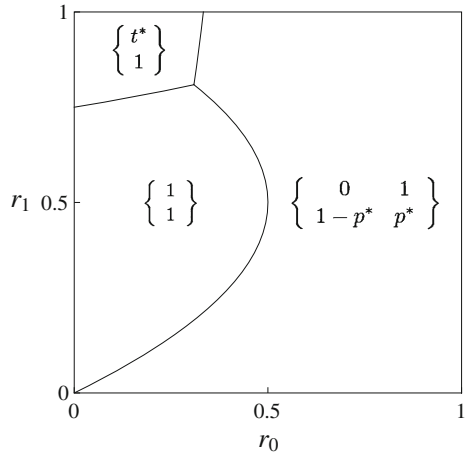
### 4 D-Optimal designs for a continuous design space

As mentioned above a design space with two or a few possible values is a first step of the analysis and it is natural to extend it to a continuous design space. In other cases, the design space is naturally continuous and we cannot simplify it. For example, in the medical example considered, the exercise time may be any value between 12 and 18 min and in the oil pricing example the price may be any value between two extremal prices. The aim of this section is to consider continuous design spaces. However, using Caratheodory’s Theorem we can restrict the search to finite designs, that is, with a finite design support. Thus, first of all we check whether the optimal designs for the two-point design space remain optimal for the continuous design space.

#### 4.1 CR D-optimal designs

When the design space is extended from  $\chi = \{0, 1\}$  to  $\chi = [0, 1]$  the D-optimal designs obtained may not remain optimal. To check if the two CR D-optimal designs

**Fig. 4** CR D-optimal designs for  $\chi = [0, 1]$



obtained are optimal for  $\chi = [0, 1]$  the GET for CR optimality (López-Fidalgo and Garcet-Rodríguez 2004, Theorem 4) is needed. For CR D-optimality the condition to be checked in the case considered here is

$$\max_{x \in [0,1]} \left[ (x, 0)M_{CR}(\theta_1, \xi^*)^{-1} \begin{pmatrix} x \\ 0 \end{pmatrix} (1 - r_x) + (x, 1)M_{CR}(\theta_1, \xi^*)^{-1} \begin{pmatrix} x \\ 1 \end{pmatrix} r_x \right] = 2.$$

Assuming  $r_x$  linear we found that the two previous CR D-optimal designs for  $\chi = \{0, 1\}$  are CR D-optimal designs for  $\chi = [0, 1]$ , except in the upper left region of Fig. 4, i.e.,

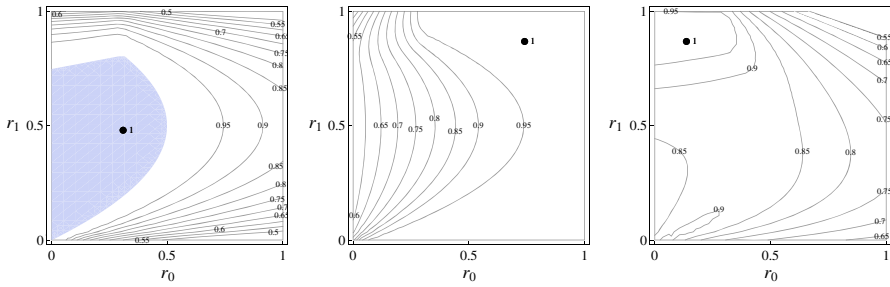
$$\xi_{CR}^*(x) = \begin{cases} \begin{Bmatrix} 0 & 1 \\ 1 - p^* & p^* \end{Bmatrix} & r_0 > 2r_1(1 - r_1) \text{ and } r_0 > \frac{r_1}{(1 + 2r_1)}, \\ \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} & \frac{r_1(-3 + 4r_1)}{-1 + 2r_1} \leq r_0 \leq 2r_1(1 - r_1), \\ \begin{Bmatrix} t^* \\ 1 \end{Bmatrix} & \frac{r_1(-3 + 4r_1)}{-1 + 2r_1} > r_0 < \frac{r_1}{(1 + 2r_1)}, \end{cases}$$

where

$$p^* = \frac{r_0}{2[r_0 - r_1(1 - r_1)]}, \quad t^* = \frac{-3 + 6r_0 + \sqrt{9 + 4(-1 + r_0)r_0}}{8(r_0 - r_1)} \in (0, 1).$$

The linear assumption of  $r_x$  would be inappropriate in some cases. In the Sect. 5 we consider a non linear case.

Now, we study the robustness of the CR D-optimal designs for  $\chi = [0, 1]$  by calculating the corresponding efficiencies using (3) as in the previous section. We consider three different nominal values  $(r_0^0, r_1^0) = \theta^0$ , one for each type of CR D-optimal design obtained for  $\chi = [0, 1]$ . The relative efficiencies are shown in Fig. 5.



**Fig. 5** Efficiencies of the CR D-optimal designs for  $\chi = [0, 1]$  with nominal parameters  $(r_0^0, r_1^0) = (1/3, 1/2)$  (left),  $(r_0^0, r_1^0) = (3/4, 7/8)$  (center) and  $(r_0^0, r_1^0) = (1/6, 7/8)$  (right) for all possible true values of  $r_0$  and  $r_1$

For the nominal values used previously in the study of the robustness of CR D-optimal designs for  $\chi = \{0, 1\}$ , that is,  $(3/4, 7/8)$  and  $(1/3, 1/2)$ , we can observe almost the same behaviour in Figs. 2 and 5. There is a small difference in the region of lower values of  $r_0$  and higher values of  $r_1$ . This is the region of the third CR D-optimal design for  $\chi = [0, 1]$ , the design concentrated at point  $t^*$ . In this area the new efficiencies are lower than before, as we can see comparing Figs. 5 and 2. When the nominal values are  $(1/6, 7/8)$  the behaviour of the relative efficiencies is quite different as we can see in the right graph of Fig. 5. Only in the small region of the third CR D-optimal design are the relative efficiencies higher. Moreover, outside this small region, the relative efficiencies are generally lower than in the other two cases. For example, when the nominal values of the parameters are  $(1/6, 7/8)$  and the true values are  $(3/4, 7/8)$  the relative efficiency (3) is 64.4 % whereas the relative efficiency is 77.8 % when we consider  $(1/6, 7/8)$  as true values and  $(3/4, 7/8)$  as nominal parameters. The same behaviour is observed in the other case: when nominal values are  $(1/6, 7/8)$  and true values are  $(1/3, 1/2)$  the relative efficiency is 86.5 %, whereas interchanging these last true and nominal values we obtain a relative efficiency of 92.2 %. Thus, if we don't know anything about the true parameters, in general, considering nominal values in the upper left region is worse than considering them elsewhere in the region.

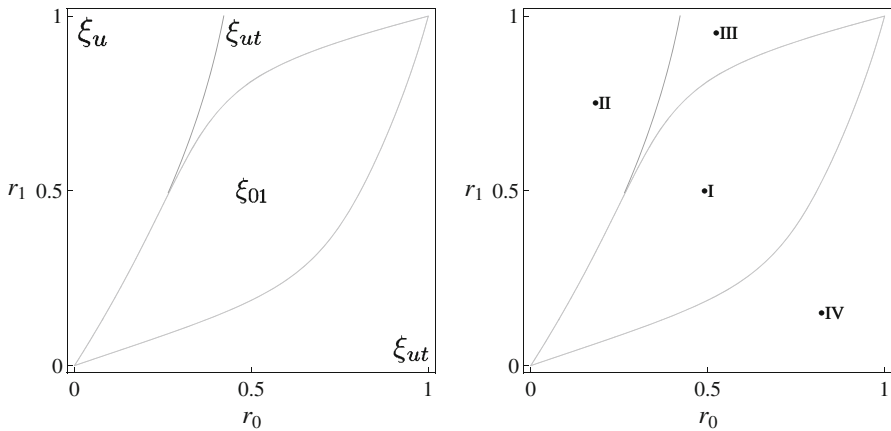
### 4.2 SE D-optimal designs

To check if the SE D-optimal design obtained previously is optimal for  $\chi = [0, 1]$  a more general version of the GET is needed (Pazman 1986, Proposition IV.26). As D-optimality is a convex criterion we can adapt the third condition of that theorem to our situation. So,  $\xi^*$  is SE D-optimal if

$$\max_{x \in \chi} \text{tr} \left[ M_{SE}^{-1}(\theta, \xi^*) I_{SE}(\theta, x) \right] = 4,$$

where  $I_{SE}(\theta, x)$  is the FIM of the design concentrated at a point  $x$ .

When a linear  $r_x$  is assumed, as in the CR case, the SE D-optimal design for  $\chi \in [0, 1]$  is SE D-optimal for  $\chi = [0, 1]$  but only in a neighbourhood of the main



**Fig. 6** SE D-optimal designs for  $\chi = [0, 1]$  assuming  $r_x$  linear (left) and nominal values of  $r_0$  and  $r_1$  considered for robustness study (right)

**Table 2** SE D-optimal designs  $\xi^*$  for nominal values  $(r_0^0, r_1^0)$  and relative efficiencies in percentages

Nominal values $(r_0^0, r_1^0)$	D-Optimal SE designs $\xi_{(r_0^0, r_1^0)}^*$	True values $(r_0^*, r_1^*)$			
		I	II	III	IV
I = $\left(\frac{1}{2}, \frac{1}{2}\right)$	$\begin{Bmatrix} 0 & 1 \\ .39 & .61 \end{Bmatrix}$	100	75	84.08	74.52
II = $\left(\frac{2}{10}, \frac{3}{4}\right)$	$\begin{Bmatrix} .718 \\ 1 \end{Bmatrix}$	84.81	100	86.44	99.05
III = $\left(\frac{11}{20}, \frac{19}{20}\right)$	$\begin{Bmatrix} .111 & .776 \\ .453 & .547 \end{Bmatrix}$	86.12	83.24	100	94.83
IV = $\left(\frac{17}{20}, \frac{3}{20}\right)$	$\begin{Bmatrix} .087 & .693 \\ .084 & .916 \end{Bmatrix}$	85.48	96.67	93.09	100

diagonal [see Fig. 6 (left)]. In the left region, for small values of  $r_0$ , the SE D-optimal design has only one support point between 0 and 1. In the remaining regions a SE D-optimal design with two support points between 0 and 1 is obtained. The limits of the regions can be obtained analytically as in the CR case, but the mathematical expressions are too long to be included here. The notation used in Fig. 6 (left) is:

$$\xi_{01} = \begin{Bmatrix} 0 & 1 \\ 1 - p^* & p^* \end{Bmatrix}, \quad \xi_u = \begin{Bmatrix} u^* \\ 1 \end{Bmatrix}, \quad \xi_{ut} = \begin{Bmatrix} u^* & t^* \\ 1 - p^* & p^* \end{Bmatrix},$$

where  $u^*, t^* \in (0, 1)$ .

As mentioned above it is interesting to check the robustness of these SE D-optimal designs for  $\chi = [0, 1]$ . We consider one representative point (parameter values) of each region for the robustness analysis; see Fig. 6 (right). Table 2 shows the nominal and true values of the parameters considered,  $\theta^0 = (r_0^0, r_1^0)$  and  $\theta^* = (r_0^*, r_1^*)$ ; the

SE D-optimal designs  $\xi_{(r_0^0, r_1^0)}^*$  and the relative efficiencies obtained using (3). We can see three SE D-optimal designs concentrated on two points and one SE D-optimal design concentrated at one point. Designs II and IV are highly efficient with respect to each other: 99.05 % when nominal values of type II and true values of type IV are considered, and 96.67 % when nominal values of type IV and true values of the parameters of type II are considered. We observe that their design weights are ‘similar’. This may be the reason why they are highly efficient with respect to each other, although they have different design types. Designs III and IV are highly efficient with respect to each other too (94.83 % and 93.09 %). In this case they are designs of the same type,  $\xi_{ut}$ , but the support points and especially the design weights are different. Design I has the lowest relative efficiencies of the entire table, 74.52 % and 75 % when nominal values of type I and true values of type IV and type II are considered respectively. Finally the relative efficiencies are also low when we consider any of the other three values as nominal values assuming (1/2, 1/2) as true values. Thus, if we have no information about the true parameters, considering nominal values of the parameters in the middle of the region is not a good choice. This is the opposite case to the robustness analysis of SE D-optimal designs when  $\chi = \{0, 1\}$ .

### 5 Discussion

In general CR and SE D-optimal designs are different approaches. Arguably SE approach can be considered more general than the CR approach. The difference is that in the SE approach for the uncontrollable variable a partially known probability distribution is considered, through estimable parameters  $\theta_2$ , while, in the CR approach this distribution is considered as a known design. If this distribution is completely known in the SE case, see Theorem 2, the D-optimal design is the same for both the CR and SE approaches. There are real problems, as oil pricing experiments (Conlisk 1979), where only a SE model can be used. A non-triangular structure deserves interesting future work.

In both approaches the D-optimal designs depend on the nominal values of the parameters. We derived formulae to obtain the D-optimal designs for a two-point design space for the model used to illustrate CR and SE methods. These designs are symmetric with respect to the parameter  $r_1$ . We conclude that the SE approach is quite robust in this context.

For the continuous design space  $[0, 1]$  we obtained the formula of the CR D-optimal design and we showed in Fig. 6 (left) the regions of the SE D-optimal design. It should be emphasized that the CR and SE D-optimal designs for  $\chi = [0, 1]$  are based on the strong assumption of a linear  $r_x$ , the probability of  $z = 1$  given the value of  $x$ . If, for example, we assume a nonlinear function  $r_x = \sin(10x) + [r_1 - r_0 - \sin(10)]x + r_0$ , the CR D-optimal designs for  $\chi = \{0, 1\}$  are not optimal for  $\chi = [0, 1]$ .

Furthermore, the linear form of  $r_x$  does not guarantee that  $0 < r_x < 1$ , thus a logistic structure may be more appropriate here. Let  $r_x$  be a logistic function of the form  $1/(1 + e^{-(\lambda + \delta x)}) \in (0, 1)$  with parameters  $\theta_2 = (\lambda, \delta)^T \in \mathbb{R}^2$ , unknown and estimable. The FIM is

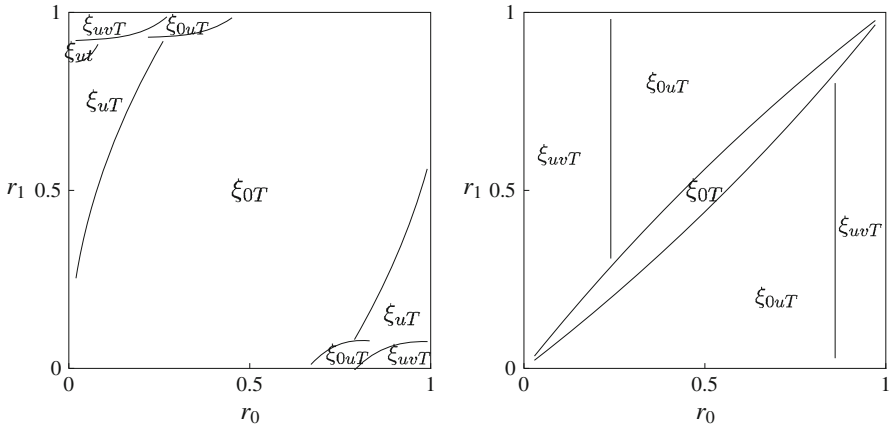


Fig. 7 SE D-optimal designs for  $\chi = [0, T]$  when  $T = 1$  (left) and  $T = 100$  (right)

$$I_{SE}(\theta; x) = \begin{pmatrix} x^2 & x - \frac{x}{1+e^{\lambda+\delta x}} & 0 & 0 \\ x - \frac{x}{1+e^{\lambda+\delta x}} & \frac{1}{1+e^{-\lambda-\delta x}} & 0 & 0 \\ 0 & 0 & \frac{1}{2+2\cosh(\lambda+\delta x)} & \frac{x}{2+2\cosh(\lambda+\delta x)} \\ 0 & 0 & \frac{x}{2+2\cosh(\lambda+\delta x)} & \frac{x^2}{2+2\cosh(\lambda+\delta x)} \end{pmatrix}.$$

In this case, the SE D-optimal design for  $\chi \in \{0, 1\}$  is again SE D-optimal design for  $\chi = [0, 1]$  but only for values of  $r_0$  “close” to  $r_1$ . When the values of  $r_0$  are far from the values of  $r_1$ , there are four different SE D-optimal designs. Figure 7 (left) shows the designs according to the values of  $r_0$  and  $r_1$ .

For a general interval  $[0, T]$  and a logistic function for  $r_x$ , SE D-optimal designs can be calculated numerically. Figure 7 shows roughly the regions of the SE D-optimal designs for extreme values  $T = 1$  and 100. The notation used in Fig. 7 is

$$\begin{aligned} \xi_{0T} &= \begin{Bmatrix} 0 & T \\ 1 - p^* & p^* \end{Bmatrix}, \quad \xi_{uT} = \begin{Bmatrix} u^* & T \\ 1 - p^* & p^* \end{Bmatrix}, \quad \xi_{ut} = \begin{Bmatrix} u^* & t^* \\ 1 - p^* & p^* \end{Bmatrix}, \\ \xi_{uvT} &= \begin{Bmatrix} u^* & v^* & T \\ 1 - p^* - q^* & p^* & q^* \end{Bmatrix}, \quad \xi_{0uT} = \begin{Bmatrix} 0 & u^* & T \\ 1 - p^* - q^* & p^* & q^* \end{Bmatrix}. \end{aligned}$$

Notice that when  $T$  grows the designs tend to have three points in their support. If  $r_0 = r_1$  the SE D-optimal design has only two support points, the extreme points 0 and  $T$ .

A sensitivity analysis for  $\chi = [0, 1]$  made in the same way as in Sect. 4.2 leads to acceptable efficiencies. It is remarkable that, again, SE D-optimal designs with nominal values for the parameters  $(r_0^0, r_1^0)$  are highly efficient with respect to SE D-optimal designs with true values  $(r_0^*, r_1^*)$  if the design weights of both designs are similar. If there is not much information about the true parameters, considering nominal values of the parameters in the upper left or lower right (by symmetry) regions is worse than considering them elsewhere in the region.

The identifiability of the structural parameters through the reduced form is an essential issue here. In the example this condition is satisfied for any function  $r_x$ . If an intercept is added to the first equation and  $r_x$  is linear then the intercept and the coefficient of  $z$  are not identifiable. Nevertheless, a non-linear  $r_x$  may help to guarantee it.

Another issue, mentioned in Sect. 2.2, is the difference between RF and RF2 models when a non linear  $\eta_{12}$  is considered. For example, let the conditional distribution of  $y|z$  be a normal distribution with mean  $\beta e^z + \gamma x$  and variance  $\sigma^2$ . Let the marginal distribution of  $z$  be a normal distribution with mean  $\alpha x$  and variance  $\tau^2$ . The equation for  $z$  given by the expectation is the same for both models, RF and RF2. For the RF model,  $y = \beta \left( e^{\alpha x + \frac{\tau^2}{2}} \right) + \gamma x + \varepsilon_1$ , taking into account the expectation of  $e^z$  is the expectation of a Log-Normal distribution. The log-likelihood is obtained as the product of the conditional density of  $y|z$  and the marginal density of  $z$ . On the other hand  $y = \beta e^{\alpha x} + \gamma x + \varepsilon_3$  for the RF2 model (plugging directly  $\alpha x$  in  $z$ ). In this case a distribution for  $(y, z)$  should be assumed. Let us consider the normal bivariate distribution with a covariance matrix obtained from the structural setup. Computing the FIM in the usual way, SE D-optimal designs were obtained for both models for  $\chi = \{0, 1\}$  and nominal values  $\alpha = -2$ ,  $\beta = 6$  (they do not depend on  $\gamma$ ) and  $\sigma = \tau = 1$ , in particular  $\xi_{RF}(1) = 0.674$  and  $\xi_{RF2}(1) = 0.787$ . The relative efficiencies are  $\text{eff}_{RF}(\xi_{RF2}) = 0.9064$  and  $\text{eff}_{RF2}(\xi_{RF}) = 0.9441$ . For different nominal values we observe that  $\text{eff}_{RF}(\xi_{RF2}) < \text{eff}_{RF2}(\xi_{RF})$ .

**Acknowledgments** We would like to thank the Editor, two anonymous reviewers and Prof. D. J. Pedregal for their very useful comments, which have improved the original version of the paper. This work has been supported by Ministerio de Educación y Ciencia and Fondos FEDER MTM2010-20774-C03-01 and Junta de Comunidades de Castilla la Mancha PEII10-0291-1850.

## References

- Aigner DJ (1979) A brief introduction to the methodology of optimal experimental design. *J Econom* 11(1):7–26
- Aigner DJ, Balestra P (1988) Optimal experimental design for error components models. *Econometrica* 56(4):955–971
- Angrist JD, Pischke J-S (2009) Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton
- Atkinson AC, Donev AN, Tobias R (2007) Optimum experimental designs, with SAS. Oxford University Press, Oxford
- Biswas A, López-Fidalgo J (2013) Compound designs for dose-finding in the presence of nondesignable covariates. *Pharm Stat* 12(2):92–101
- Conlisk J (1979) Design for simultaneous equations. *J Econom* 11(1):63–76
- Draper NR, Pukelsheim F (1996) An overview of design of experiments. *Stat Pap* 37(1):1–32
- Hahn J, Keisuke H, Dean K (2011) Adaptive experimental design using the propensity score. *J Bus Econ Stat* 29(1):96–108
- Kessels R, Jones B, Goos P, Vandebroek M (2009) An efficient algorithm for constructing Bayesian optimal choice designs. *J Bus Econ Stat* 27(2):279–291
- Kiefer J (1974) General equivalence theory for optimum designs (approximate theory). *Ann Stat* 5:849–879
- López-Fidalgo J, Garcet-Rodríguez S (2004) Optimal experimental designs when some independent variables are not subject to control. *J Am Stat Assoc* 99:1190–1199
- Martín-Martín R, Torsney B, López-Fidalgo J (2007) Construction of marginally and conditionally restricted designs using multiplicative algorithms. *Comput Stat Data Anal* 51:5547–5561

- Newey WK, Powell JL, Vella F (1999) Nonparametric estimation of triangular simultaneous equations models. *Econometrica* 67(3):565–603
- Pazman A (1986) Foundations of optimum experimental design. D. Reidel Publishing Company, Dordrecht

# Robust integral compounding criteria for trend and correlation structures

M. Stehlík · J. López-Fidalgo · V. Casero-Alonso ·  
Elena Bukina

© Springer-Verlag Berlin Heidelberg 2014

**Abstract** Optimal design is a crucial issue in Environmental measurement with typical time–space correlated observations. A modified Arrhenius model with a particular correlation structure will be applied to the methane removal in the atmosphere, a very important environmental issue at this moment. We introduce a class of integrated compound criteria for obtaining robust designs. In particular, the paper provides an insight into the relationship of a compound D-optimality criterion for both the trend and covariance parameters, and the Integrated Mean Squared Prediction Error (IMSPE) criterion. In general, if there are two or more approaches of a given problem, e.g. two rival models or two different parts of a model, an integral relationship may be constructed with the aim of finding a suitable compromise between them. The Fisher information matrix (FIM) will be used in both cases. Then the

integral compound criterion with respect to a density from a given parametric family of distributions is optimized. We also discuss some general conditions around the behavior of the introduced approach for comparing the FIMs and provide computing methods.

**Keywords** Correlated errors · Efficiency · Equidistant design · Experimental design · Fredholm equation · Parameterized covariance functions · Regularization.

## 1 Introduction

Comparison of FIMs from different models have been considered in different ways in the statistical literature [see e.g. Ahmadi and Arghami (2003), Hofmann (2004) or Alshunnar et al. (2012)]. If there are two or more approaches of a given problem, e.g. two rival models or two different parts of a model, an integral relationship may be constructed with the aim of finding a suitable compromise between them. The purpose of this is to obtain robust designs for both matrices. The usual compound criteria may not be powerful enough to do this in a proper way.

Observations from various environmental measurements are often approximated as realizations of correlated random fields. Such an approach is interesting for assessing e.g. drought and flood risks [see e.g. Unami et al. (2010)]. The robust designs obtained in the current paper can be of interest for irregular sampling, studied e.g. in Tandeo et al. (2011) for the case of aggregation of many meteorological and oceanographic variables from satellites. Rodríguez-Díaz et al. (2012) gave optimal designs for a modified Arrhenius model, used for modeling a flux of methane in troposphere. This has an important impact in greenhouse gas emission. Kinetics of chemical reactions are usually

---

M. Stehlík (✉)

Departamento de Matemática, Universidad Técnica Federico  
Santa María, Casilla 110-V, Valparaíso, Chile  
e-mail: mlnstehlik@gmail.com; milan.stehlik@usm.cl

M. Stehlík

Department of Applied Statistics, Johannes Kepler University in  
Linz, Linz, Austria  
e-mail: milan.stehlik@jku.at

J. López-Fidalgo · V. Casero-Alonso

Department of Mathematics, University of Castilla-La Mancha,  
Ciudad Real, Spain  
e-mail: jesus.lopezfidalgo@uclm.es

V. Casero-Alonso

e-mail: victormanuel.casero@uclm.es

E. Bukina

Laboratoire I3S, CNRS/Université de Nice-Sophia Antipolis,  
Sophia Antipolis, France  
e-mail: bukina@i3s.unice.fr

considered at the Earth. Thus, this problem provides an unusual environment for kinetic reactions, where the correlation structure and the modified Arrhenius equation can play a crucial role. In particular, it was applied to the methane removal in the atmosphere, a very important issue at this moment. This case is used in Sect. 5.1 to illustrate the method.

In a model with correlated observations, where the trend and the covariance structure have different parameters the FIM is a block diagonal matrix. Putting both matrices at the same level may not be appropriate at all. The method proposed in this paper deals with this situation producing robust designs for estimating both parts of the model. In this Section a motivating example is given where the usual compound criterion never pays attention to one part of it. In order to introduce properly the motivating example some background on optimal experimental design is given in the next Section.

### 1.1 Optimal experimental design background

Let  $y$  be the response variable and assume  $x$  is a vector of explanatory variables. Assume the experiment is realized  $n$  times for  $n$  values,  $x_1, x_2, \dots, x_n$ , of  $x$ . Let  $Y_1(x_1) \equiv y_1, Y_2(x_2) \equiv y_2, \dots, Y_n(x_n) \equiv y_n$  be the corresponding outcomes. Let be a general model describing the relationship between both groups of variables through a parametric family of distributions defined by the probability density functions (pdf's),

$$\{h(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m) | \theta = (\theta_1, \theta_2, \dots, \theta_m)^T \in \Theta\}.$$

An experimental condition  $x$  can be chosen on a compact design space,  $\chi$ . An exact design of size  $n$  is defined by  $\xi = (x_1, x_2, \dots, x_n)$ , where some of them may be repeated. Thus, a probability measure can be defined with support on the different points and weights proportional to the number of repetitions, say  $\xi(x)$  for each  $x \in \chi$ . The FIM, under typical regularity assumptions, for a particular experimental condition  $x$  is

$$I(x, \theta) = E_h \left( \frac{\partial L(x, \theta)}{\partial \theta} \frac{\partial L(x, \theta)}{\partial \theta^T} \right) = - E_h \left( \frac{\partial^2 L(x, \theta)}{\partial \theta^2} \right).$$

The expectation  $E_h$  is computed here with respect to the distribution with pdf  $h(y|x; \theta)$  and  $L(x, \theta)$  is the log-likelihood for this distribution. This matrix is a measure of the information provided by  $x$  to the estimation of the model. The associated FIM will be

$$M(\theta) = \sum_{x \in \chi} \xi(x) I(x, \theta),$$

where for simplicity of notation the dependence of the FIM on an experimental design,  $\xi$ , will be omitted in the

notation hereafter. A popular criterion is D-optimality, based on the maximization of the determinant of the information matrix, that is minimization of the determinant of the covariance function of the estimators of the parameters, at least asymptotically under some conditions [see e.g. Pázman (2010)].

The existence of error correlation structure introduces a number of issues, which require special techniques in the traditional optimal design theory [for a recent discussion see Müller and Stehlík (2009)]. The statistical model we consider in this section is a random field

$$Y(x) = \eta(x, \theta_1) + \varepsilon_{\theta_2}(x)$$

with design points (coordinates of monitoring sites)  $\xi_n = (x_1, \dots, x_n)$  taken from a compact design space  $\mathcal{X} = X^n, X = [a, b], -\infty < a < b < \infty$ . The parameters  $\theta_1$  are unknown and the variance-covariance structure of the errors depends on parameters  $\theta_2$ . However, if distributional assumptions are made, one can employ the Maximum Likelihood Estimators. If there is not overlapping between  $\theta_1$  and  $\theta_2$  the information matrix exhibits a block diagonal form

$$\begin{pmatrix} M_1(\theta) & 0 \\ 0 & M_2(\theta) \end{pmatrix}.$$

Outputs from various environmental measurements are often approximated as realizations of correlated random fields. Two approaches are considered to design experiments for a correlated random field when the objective is to obtain precise predictions over the whole experimental domain. The first one corresponds to a compound  $D_\alpha$ -optimality criterion for both the trend and the covariance parameters introduced by Müller and Stehlík (2010),

$$\phi_{1*}(\theta, M_1(\theta), M_2(\theta)) = |M_1(\theta)|^{\theta_3} |M_2(\theta)|^{1-\theta_3}, \quad 0 \leq \theta_3 \leq 1,$$

where  $\theta_3$  plays here the role of  $\alpha$  and  $M_1(\theta)$  and  $M_2(\theta)$  are the FIMs for the parameters  $\theta_1$  and  $\theta_2$ .

The second one relies on an approximation of the mean squared prediction error already proposed in the literature. In particular, Müller and Pronzato (2009) conjectured and showed on an example that for some particular settings both approaches yield similar optimal designs, thereby revealing a sort of equivalence theorem for random fields. For estimations of spatial fields a classical criterion is the Empirical Kriging (EK) prediction error. Here we have to minimize the so-called Kriging variance  $\text{Var}[\hat{Y}(x|\xi)] = E[(\hat{Y}(x|\xi) - Y(x))^2] \equiv \text{MSPE}(x|\xi)$  (Mean Squared Prediction Error), where  $\hat{Y}(x|\xi)$  denotes the best linear unbiased predictor (BLUP) of  $Y(x)$  based on a particular design  $\xi$ . Thus, the EK-optimal design minimizes the criterion function

$$\psi(\xi) = \max_{x \in X} \text{Var}[\widehat{Y}(x|\xi)].$$

However, this criterion is difficult to manage. A much easier criterion is  $D_\alpha$ -optimality.

### 1.2 Integral compounding

Let consider two models, which may differ in their statistical representation. Then we can construct two different FIMs. This is the case of the trend and the covariance models. Thus, let  $M_1(\theta), M_2(\theta)$  be two FIMs with  $\theta \in \Theta$ . Let  $u : \Theta \rightarrow R$  be a function from a set  $D$ , eventually a parametric family of functions, e.g. pdf's. We will consider a compound integral criterion

$$Lu = \int_{\Theta} \phi(\theta, M_1(\theta), M_2(\theta))u(\theta)d\theta, \tag{1}$$

where  $\phi$  is a particular kernel. It is important to mention that the parameter vector here includes model parameters and other parameters or constants related to the procedure, e.g. new parameters defining the kernel.

The aim is to find an appropriate function  $u^*$ , e.g. maximizing or minimizing  $Lu$ , or solving the Fredholm integral equation  $Lu = \psi$ , where  $\psi$  is a given criterion. Then a design will be constructed from this function. This will provide a suitable robust optimal design. Notice, that optimizing  $Lu$  may be different than optimizing its kernel  $\phi$ .

Typical examples for the kernel,  $\phi$ , are

1.  $\phi_1(\theta, M_1(\theta), M_2(\theta)) = |M_1(\theta)|^{\theta_3/m_1} |M_2(\theta)|^{(1-\theta_3)/m_2}$ .
2.  $\phi_2(\theta, M_1(\theta), M_2(\theta)) = \frac{\theta_3}{m_1} \log |M_1(\theta)| + \frac{1-\theta_3}{m_2} \log |M_2(\theta)|$ .
3.  $\phi_{2^*}(\theta, M_1(\theta), M_2(\theta)) = \theta_3 \log |M_1(\theta)| + (1 - \theta_3) \log |M_2(\theta)|$ .
4.  $\phi_3(\theta, M_1(\theta), M_2(\theta)) = \theta_3 \left( \frac{|M_1(\theta)|}{|M_1^*(\theta)|} \right)^{1/m_1} + (1 - \theta_3) \left( \frac{|M_2(\theta)|}{|M_2^*(\theta)|} \right)^{1/m_2}$ , where  $|M_i^*(\theta)|$  corresponds to the maximum determinant of  $|M_i(\theta)|$ ,  $i = 1, 2$ .
5.  $\phi_{3^*}(\theta, M_1(\theta), M_2(\theta)) = \theta_3 \left( \frac{|M_1(\theta)|}{|M_1^*(\theta)|} \right) + (1 - \theta_3) \left( \frac{|M_2(\theta)|}{|M_2^*(\theta)|} \right)$ .

Here  $|\cdot|$  stands either for the determinant or the absolute value,  $\theta^T = (\theta_1, \theta_2, \theta_3)$ ,  $0 \leq \theta_3 \leq 1$  and  $m_i$  is the dimension of sub-vector  $\theta_i$  of  $\theta$ ,  $i = 1, 2$ . After finding an appropriate function  $u$  an optimal design is searched by other means as minimizing or maximizing  $Lu(\xi)$ , or else constructing the design using a function  $u$ .

The function  $u$  may be fixed from the beginning if there is some interest or knowledge on the parameters of this problem. Trivial examples are the Dirac measure, which leads to locally optimizing the kernel, or  $u = 1$ , which

takes to optimize the simple average of the kernel on the parameters. Other typical distributions are the uniform or the Beta family.

As the following results state some kernels are equivalent, that is, they produce the same spectrum of optimal designs. Let  $\phi$  and  $\phi'$  two concave optimality criteria defined on an information matrices space, let  $\lambda \in [0, 1]$  and  $\phi_\lambda = \lambda\phi + (1 - \lambda)\phi'$ . The design maximizing  $\phi_\lambda$  will be called  $\xi_\lambda$ .

**Lemma 1** Let  $\varphi : \mathfrak{R} \rightarrow \mathfrak{R}$  be a continuous increasing function. Then the criteria  $\tilde{\phi}_\mu = \mu\phi + (1 - \mu)\varphi \circ \phi'$  and  $\phi_\lambda$  are equivalent in the sense that for any  $\lambda \in [0, 1]$  there exists  $\mu \in [0, 1]$  such that  $\xi_\lambda = \tilde{\xi}_\mu$ , where the later is a design maximizing  $\tilde{\phi}_\mu$ .

*Proof* Since  $\phi'$  is concave and  $\varphi$  is increasing then  $\varphi \circ \phi'$  is concave. From Cook and Wong (1994) given  $\lambda \in [0, 1]$  there exist a constant  $C_\lambda$  such that  $\xi_\lambda = \arg \max \{\phi(\xi) | \phi'(\xi) \geq C_\lambda\}$ . But  $\phi'(\xi) \geq C_\lambda$  if and only if  $\varphi[\phi'(\xi)] \geq \varphi(C_\lambda)$ . Applying again the results of Cook and Wong (1994) to  $\tilde{\phi}_\mu$ , given  $\varphi(C_\lambda)$  there exists  $\mu \in [0, 1]$  such that  $\xi_\lambda = \arg \max_\xi \tilde{\phi}_\mu(\xi) = \tilde{\xi}_\mu$ .  $\square$

### Corollary 1

1. Let  $\varphi, \varphi' : \mathfrak{R} \rightarrow \mathfrak{R}$  continuous and increasing then the criteria  $\tilde{\phi}_\mu = \mu\varphi \circ \phi + (1 - \mu)\varphi' \circ \phi'$  and  $\phi_\lambda$  are equivalent.
2. Criterion  $\tilde{\phi}_\mu = \mu^\mu \phi^{1-\mu}$  is equivalent to  $\tilde{\phi}_\mu = \mu \log \phi + (1 - \mu) \log \phi'$ , which is equivalent to  $\phi_\lambda$ .

This means in particular that the same optimal design is obtained for the two criteria for different values  $\lambda$  and  $\mu$  in each criterion.

If the function  $u(\theta)$  does not depend on the parameters of the model(s) then we get the traditional compound design as shown in the following Lemma.

**Lemma 2** If  $u(\theta)$  depends only on  $\lambda$  then  $Lu = \int_0^1 [\lambda\phi + (1 - \lambda)\phi']u(\lambda)d\lambda$  is equivalent to the compound criterion  $\phi_\mu = \mu\phi + (1 - \mu)\phi'$  with  $\mu = \frac{\int_0^1 \lambda u(\lambda)d\lambda}{\int_0^1 u(\lambda)d\lambda}$ .

### 1.3 Motivating example: misspecification of linear and exponential correlation

Let us consider the model  $Y(x_i) = \varepsilon(x_i), x_i \in [0, 1]$  where errors are correlated according to either an exponential,  $\exp(-\theta d)$ , or a linear,  $1 - \theta d$ , covariance function. Here  $d = |x_i - x_j|$ , and  $\theta > 0$  is an unknown parameter of interest. Notice that  $\exp(-\theta d) = 1 - \theta d + o(\theta d)$ , where  $o$

is the “little-o” Landau symbol defined as follows:  $f(x) = o(g(x))$  means that  $f/g \rightarrow 0$  as  $x \rightarrow 0$ . For simplicity, we consider exact two-point optimal designs,  $(x, x + d)$ , for the covariance parameter  $\theta$ . Typical approaches in case of covariance uncertainty consider individual models and compounding. In the next a, b, c we show how integrated compound criteria may regularize collapsing of the design points. By collapsing we understand that the criterion attains its maximum at  $d = 0$ .

- (a) *Optimal design for exponential correlation* The FIM for  $\theta$  under exponential correlation has the form  $M_{exp}(\theta) = \frac{d^2 \exp(-2\theta d)(1 + \exp(-2\theta d))}{(1 - \exp(-2\theta d))^2}$ . The maximum is obtained for  $d = 0$ , i.e. the optimal design collapses.
- (b) *Optimal design for linear correlation* This covariance structure is defined by a linear semi-variogram  $\gamma(d) \equiv \frac{1}{2} \text{var}(Y(x) - Y(x + d)) = \theta d$  with zero nugget, that is, there is not additional term for the variance. For regularity we assume that  $\theta d \leq 2, \theta > 0$  (thus, negative correlation is allowed; to avoid this, we need  $\theta d < 1$ ). Then the maximal FIM is obtained for maximal  $d$ . To verify this, we obtain the FIM

$$M_{lin}(\theta) = \frac{-2\theta d + \theta^2 d^2 + 2}{\theta^2(\theta d - 2)^2}.$$

We have

$$\frac{\partial M_{lin}(\theta)}{\partial d} = \frac{2d}{(2 - \theta d)^3}.$$

So  $M_{lin}(\theta)$  is an increasing function for every (acceptable)  $0 < d < \min\{2/\theta, 1\}$ , otherwise it would be decreasing. Thus  $d^* = \min\{2/\theta, 1\}$ .

- (c) *Compound criterion and its regularization by an integral compound criterion* The criterion  $\theta \log M_{exp}(\theta) + (1 - \theta) \log M_{lin}(\theta)$  attains the maximum for  $d = 0$

(collapsing again) for any  $\theta \in [0, 1]$ . In the situation when the exponential covariance is the true one this design will have only about 15 % of efficiency.

Now we will illustrate the potential of regularized integral compound criteria. Taking kernel

$$\phi_{1^*}(\theta, M_{exp}(\theta), M_{lin}(\theta)) = |M_{exp}(\theta)|^\theta |M_{lin}(\theta)|^{1-\theta},$$

the integral compound criterion has the form

$$Lu = \int_0^1 \phi_{1^*}(\theta, M_{exp}(\theta), M_{lin}(\theta)) u(\theta) d\theta.$$

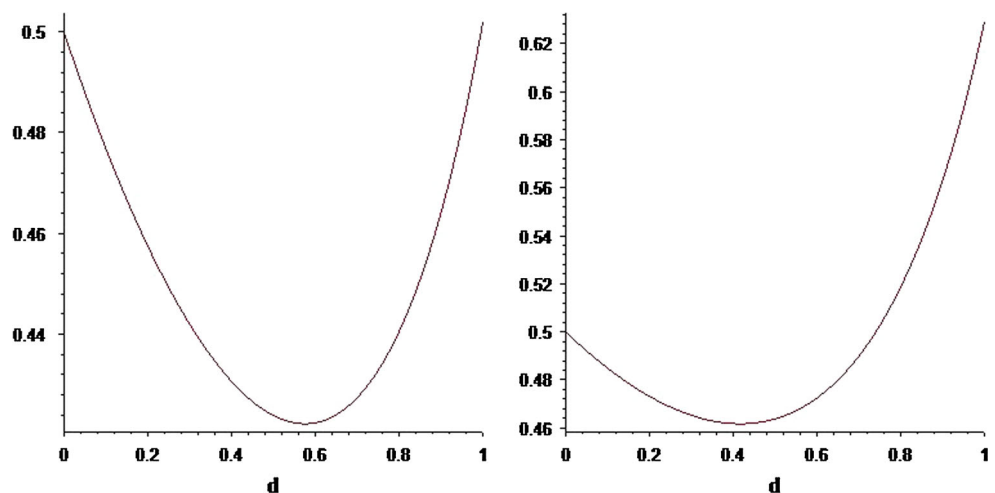
Let us take  $u^*(\theta) = 1$  (i.e. the density of the uniform distribution over  $[0, 1]$ ). Then maximum of  $Lu^*$  for  $\theta = 1$  is received at  $d^* = 1$ , i.e. it is not collapsing and it has efficiency 100 % for the linear covariance on the design space  $[0, 1]$ , as we can see in Fig. 1 (left). If we use non-uniform prior on parameter  $\theta$ , e.g. having density  $2\theta$ , the effect is even more significant as we can see in Fig. 1 (right).

Detailed proofs of the assertions above are included in the “Appendix”.

### 1.4 Structure of the paper

The paper is organized as follows. In Sect. 2 we study some properties of the class of integrated compound criteria for IMSPE approximation. We provide (by means of Fredholm integral operators) the theoretical background for existence of specific criteria, related to IMSPE. In Sect. 3 theoretical arguments are provided for the need of the regularization of the Fredholm equation. In Sect. 4 we illustrate the methodology by numerical experiments. This approach is applied to the simultaneous estimation. In Sect. 5 we compare the FIM for both cases of simultaneous equations and conditionally restricted designs and we optimize the functional (1) with respect to the density from a given

**Fig. 1** Information gain for two integrated compound designs



parametric family (namely Beta). This example considers two different approaches to a model with an explanatory variable that is not under the control of the experimenter, but it can only be observed after the experiment is realized. The two approaches are simultaneous equations models on the one hand and conditionally restricted designs on the other [Casero-Alonso and López-Fidalgo (2014), López-Fidalgo and Garcet-Rodríguez (2004)]. A Discussion section concludes the paper.

## 2 Compound criteria for IMSPE

The main aim of the kriging technique consists in predicting output of a simulator on the experimental region, and for any untried location  $x$  the estimation procedure is focused on the BLUP,  $\hat{Y}(x|\xi)$ . Thus natural criteria will minimize suitable functionals of the MSPE. Since often the prediction accuracy is related to the entire prediction region  $\chi = X^n$ , a very practical design criterion is the Integrated MSPE given by  $IMSPE(x|\xi) = \int_{X^n} \sigma^{-2} MSPE(x|\xi) d\xi(x)$ . IMSPE was used in several papers [see Sacks et al. (1989), Crary (2002)] for construction of optimal designs.

Some of the properties of criterion (1) for  $\phi_{1^*}$  will be studied in this paper, which is given by the direct integration of  $D_x$ -optimality criterion. There are several possible extensions. One possible modification is related to the region for  $\theta$ . There is no particular reason for restriction  $\theta \in [0, 1]$ . However, we assume in this paper that  $\theta$  belongs to a compact interval and thus ‘for the sake of simplicity’ we normed this interval to be  $[0, 1]$ . An important question is the dimensionality of  $\theta$ . For designs with large number of points a multivariate function  $u(\theta_3, \dots, \theta_m)$  may be more appropriate. Then

$$Lu = \int_{[0,1]^{m-2}} |M_1(\theta)|^{\theta_3} \dots |M_1(\theta)|^{\theta_m} |M_2(\theta)|^{1-\theta_3} \dots |M_2(\theta)|^{1-\theta_m} u(\theta_3, \dots, \theta_m) d\theta_3 \dots d\theta_m.$$

An alternative integrated criterion is given by  $\phi_1$  and the Fredholm equation  $\psi = Lu$ . From numerical reasons (ill-conditioned matrix) we use the criterion  $\phi_{1^*}(\theta_3, M_1(\theta), M_2(\theta)) = \frac{\theta_3}{m_1} \log |M_1(\theta)| + \frac{1-\theta_3}{m_2} \log |M_2(\theta)|$ . Notice, that in the latter case as stated in Lemma 2 we obtain a criterion of the form  $\frac{A}{m_1} \log |M_1(\theta)| + \frac{1-A}{m_2} \log |M_2(\theta)|$ , where  $A = \int_0^1 \theta_3 u(\theta_3) d\theta_3$ , i.e.  $A$  is the mean of the random variable  $\theta_3$  (since from practical reasons we later will regularize it by  $u \geq 0, \|u\| = 1$ ). Moreover, this approach relates to a general compound criterion treated in the literature [see e.g. McGree et al. (1988)].

### 2.1 Integrated compound criteria mimics the nugget effect

As was seen in the motivating example in the Introduction a non-constant  $u$  may regularize a ‘‘collapsing’’ effect [see e.g. Kiseřák and Stehlík (2008)]. Thus, it may be said that integrating  $\phi_1(\theta, M_1(\theta), M_2(\theta))$  with non-constant  $u$  works, in this sense, similarly as the so called ‘‘nugget effect’’.

Let us consider an Ornstein–Uhlenbeck (OU) process for a 2-point design where the estimation of the correlation parameter is of interest,

$$E [Y(x)] = \theta_1, \text{ corr } (Y(x), Y(x + d)) = \exp(-\theta_2 d), x \in [0, 1].$$

The information matrix for both parameters is diagonal with these two elements in it, apart from the variance,

$$M_1(\theta) = \frac{2 \exp(\theta_2 d)}{1 + \exp(\theta_2 d)} \text{ and } M_2(\theta) = \frac{d^2(1 + \exp(2\theta_2 d))}{(1 - \exp(2\theta_2 d))^2}.$$

The kernel  $\phi_1$  is considered here with a nominal value  $\theta_2 = 1$ . If two-point exact designs are searched for each value of the parameters maximizing the kernel, it can be seen that the optimal value for the distance is

$$d^* = \begin{cases} 0 & \theta_3 \leq 2/3, \\ \in (0, 1) & 2/3 < \theta_3 < 0.7627, \\ 1 & \theta_3 \geq 0.7627. \end{cases}$$

If the tuning parameter  $\theta_3$  is assumed equal to the parameter  $\theta_2$ , say  $\theta \equiv \theta_3 = \theta_2$ , then a similar result is obtained,

$$d^* = \begin{cases} 0 & \theta \leq 2/3, \\ \in (0, 1) & 2/3 < \theta < 0.7215, \\ 1 & \theta \geq 0.7215. \end{cases}$$

In both cases for  $\theta_3 \leq 2/3$  there is the so called ‘‘collapsing effect’’, reported recently by Crary (2002) and Müller and Stehlík (2009). Let us consider the approach introduced in this paper.

#### Lemma 3

1. If  $\theta \equiv \theta_3 = \theta_2$  and  $u(\theta) = 1$  then

$$\arg \max_d (Lu)(d) = \arg \max_d \int_0^1 \phi_{1^*}(\theta, M_1(\theta), M_2(\theta)) u(\theta) d\theta = 0.$$

Thus, there is not improvement in the collapsing effect at all. Let consider other functions.

2. If  $u(\theta) = \theta, \theta^2, e^\theta, \sin \theta$  then

$$d^* = \arg \max_{d \in [0,1]} (Lu)(d) > 0.$$

In particular,  $d^* = 1$  for all and  $d^* = 1.1411, 1.5280, 1.8121, 1.0645$  respectively if the design space is large enough.

This means an important improvement since collapsing does not occur anymore. By this we have obtained the regularization of collapsing by integral compound criteria. The other way of regularizing of collapsing is to employ a nugget [Müller and Stehlík (2009)].

### 2.2 Integrated compound criteria mimics IMSPE

Let consider again the Ornstein–Uhlenbeck model with constant trend and  $n$  design points:  $x, x + d_1, x + d_1 + d_2, \dots, x + d_1 + \dots + d_{n-1}$ . Then [Baldi Antognini and Zagoraiou (2010)]

$$\psi(\xi) = -IMSPE(x|\xi) = -1 + \frac{n-1}{\theta_2} - 2A(\xi) - \frac{C(\xi)}{B(\xi)},$$

where

$$A(\xi) = \sum_{i=1}^{n-1} \frac{d_i}{\exp(2\theta_2 d_i) - 1},$$

$$B(\xi) = \sum_{i=1}^{n-1} \left[ d_i + \frac{\exp(\theta_2 d_i) - 1}{\exp(\theta_2 d_i) + 1} \right],$$

$$C(\xi) = \sum_{i=1}^{n-1} \left[ d_i + \frac{3(1 - \exp(2\theta_2 d_i)) + 2\theta_2 d_i \exp(\theta_2 d_i)}{\theta_2 (\exp(\theta_2 d_i) + 1)^2} \right],$$

and the kernel is

$$\phi_1(\theta, M_1(\theta), M_2(\theta)) = \left( 1 + \sum_{i=1}^{n-1} \frac{\exp(\theta_2 d_i) - 1}{\exp(\theta_2 d_i) + 1} \right)^{\theta_3}$$

$$\left( \sum_{i=1}^{n-1} \frac{d_i^2 (\exp(2\theta_2 d_i) + 1)}{(\exp(2\theta_2 d_i) - 1)^2} \right)^{1-\theta_3}.$$

In order to compute optimal designs some methods are adapted and evaluated in the next section.

### 3 Numerical solution of the problem

The problem of solving numerically a Fredholm equation  $Lu = \psi$  of the first kind is complicated by the fact that the inversion operator is not in general continuous so that numerical stability is a major problem. A common approach is to find a regularized solution by minimizing the functional  $\|Lu - \psi\|^2 + \lambda \Omega(u)$  over a suitable set of functions  $D$ , where  $\Omega$  is a stabilizing non-negative functional chosen to be small for  $u$  having desirable properties (typically smoothness) and  $\lambda$  is a regularization parameter giving an appropriate balance (trade-off) between the value

of  $\Omega(u)$  and the error  $\|Lu - \psi\|^2$  [see e.g. Tikhonov and Arsenin (1977) for a seminal paper]. The so-called Tikhonov regularization has been extensively studied,

$$\Omega(u) = \int \sum_{i=0}^r a_i |f^{(i)}|^2,$$

where  $a_i$  is a real constant and  $f^{(i)}$  the  $i$ -th derivative.

Wahba (1977) developed the generalized cross validation method to estimate  $\lambda$  directly from the data. Hansen (1992) stated the need of regularization in the discretized situation. The numerical experiments given in Sect. 4 illustrate the need of regularization in this setup.

In what follows the minimization is considered reformulating the problem of equivalence of two optimality criteria as a quadratic optimization problem. Numerical methods will be used to solve it. The equivalence conjecture states there exists a function  $u^*(\theta)$  that the following holds

$$\psi(\xi) \simeq \int_0^1 \phi(\theta, M_1(\theta), M_2(\theta)) u^*(\theta) d\theta.$$

Here  $\psi(\xi)$  denotes a criterion related to prediction and  $\phi(\xi|\theta) \equiv \phi(\theta, M_1(\theta), M_2(\theta))$  denotes a kernel.

The problem is how to find such a function  $u(\theta)$ , so that the two criteria are equivalent. Thus, we are interested in the minimization of the following function

$$f_\xi = \left[ \psi(\xi) - \int_0^1 \phi(\theta, M_1(\theta), M_2(\theta)) u(\theta) d\theta \right]^2,$$

with respect to  $u(\theta)$ .

For the sake of simplicity, consider that  $u(\theta)$  is of the form

$$u(\theta) = \sum_{i=1}^r w_i \delta_{\theta^{(i)}}(\theta), \tag{2}$$

where  $w = (w_1, \dots, w_r)^T$  are some non negative weights,  $\Theta = (\theta^{(1)}, \dots, \theta^{(r)})^T$  is a partition of  $\theta$  with  $0 \leq \theta^{(1)} < \theta^{(2)} \dots \theta^{(r)} \leq 1$ , and  $\delta_{\theta^{(i)}}$  is the Kronecker function. Designs  $\xi_j, j = 1, \dots, q$  are chosen at this step and then our problem can be written as

$$f = \sum_{j=1}^q f_{\xi_j} = \sum_{j=1}^q \left[ \psi(\xi_j) - \sum_{i=1}^r w_i \phi(\xi_j|\theta^{(i)}) \right]^2,$$

$$= \sum_{j=1}^q [\psi(\xi_j) - \phi^T(\xi_j|\Theta)w]^2,$$

$$= \sum_{j=1}^q \psi^2(\xi_j) + w^T \left( \sum_{j=1}^q \phi(\xi_j|\Theta) \phi^T(\xi_j|\Theta) \right) w$$

$$- 2w^T \sum_{j=1}^q \psi(\xi_j) \phi(\xi_j|\Theta),$$

Thus the problem of minimizing function  $f$  is equivalent to minimizing the following function

$$\tilde{f} = \frac{1}{2} w^T A w - b^T w + c, \tag{3}$$

where

$$A = \sum_{j=1}^q \phi(\xi_j|\Theta)\phi^T(\xi_j|\Theta), \quad b = \sum_{j=1}^q \psi(\xi_j)\phi(\xi_j|\Theta),$$

$$c = \frac{1}{2} \sum_{j=1}^q \psi^2(\xi_j).$$

Hence, our aim is to find the optimum of a quadratic function  $\tilde{f}$ . The Hessian  $A$  of the objective function  $\tilde{f}$  is a symmetric positive semi-definite matrix. Actually, for a fixed  $\xi$ ,  $\phi(\xi|\Theta)$  is a column vector

$$\phi(\xi|\Theta) = \begin{pmatrix} \phi(\xi|\theta^{(1)}) \\ \vdots \\ \phi(\xi|\theta^{(r)}) \end{pmatrix},$$

and thus  $\phi_j(\xi|\Theta)\phi_j^T(\xi|\Theta)$  is a symmetric matrix

$$\phi_j(\xi|\Theta)\phi_j^T(\xi|\Theta) = \begin{pmatrix} \phi_j^2(\xi|\theta^{(1)}) & \dots & \phi_j(\xi|\theta^{(1)})\phi_j(\xi|\theta^{(r)}) \\ \vdots & \ddots & \vdots \\ \phi_j(\xi|\theta^{(r)})\phi_j(\xi|\theta^{(1)}) & \dots & \phi_j^2(\xi|\theta^{(r)}) \end{pmatrix}.$$

To construct the matrix  $A$  we sum up to  $q$  (number of the chosen designs  $\xi$ ) matrices of this form, so we also get a symmetric matrix.

### 3.1 Constrained case

So far we considered that there are no constraints on the function  $u(\theta)$ . However, it can be useful to require  $u(\theta)$  to have some certain properties. For instance, we can impose the following constraints on  $u(\theta) \geq 0$

$$\int_0^1 u(\theta)d\theta = 1.$$

This corresponds to a constrained quadratic optimization problem. Since we assumed that  $u(\theta)$  is of the form  $u(\theta) = \sum_{i=1}^r w_i \delta_{\theta^{(i)}}$ , the constraint here can be simply rewritten as  $\sum_{i=1}^r w_i = 1$ .

The usual trick when solving a constrained quadratic optimization problem is a penalty approach. It consists in introducing some penalty term  $\sigma$  to the original function  $\tilde{f}$

and then solving a corresponding (equivalent) unconstrained problem.

In our case first we rewrite the constraint  $\sum_{i=1}^r w_i = 1$  in the form  $J^T w = 1$ , where  $J = (1, \dots, 1)^T$ , and then we define

$$f_\sigma(x) = \tilde{f} + \sigma(J^T w - 1)^2 = \tilde{f} + \sigma(w^T J J^T w - 2w^T J + 1),$$

$$= \frac{1}{2} w^T A w - w^T b + c + \sigma(w^T J J^T w - 2w^T J + 1)$$

$$= \frac{1}{2} w^T (A + 2\sigma J J^T) w - w^T (b + 2\sigma J) + \sigma + c.$$

Now we can solve this unconstrained problem with some large penalty term  $\sigma$  instead of the original constrained problem.

### 3.2 Weighted case

We also can be interested in minimization of the weighted function

$$f_\xi^a = \left[ \psi(\xi) + a - \int_0^1 \phi(\xi|\theta)u(\theta) d\theta \right]^2,$$

where  $a$  is some scaling factor. We shall minimize  $\sum_{i=1}^q f_{\xi_i}^a$  with respect to  $u(\theta) \geq 0$ , such that  $\int_0^1 u(\theta)d\theta = 1$ . Then we minimize the resulting function with respect to the scaling factor  $a$ . Thus, we have

$$\min_a \min_{u(\theta): \int_0^1 u(\theta)d\theta=1} \sum_{j=1}^q f_{\xi_j}^a.$$

Again,

$$f_a = \sum_{j=1}^q f_{\xi_j}^a = \sum_{j=1}^q \left[ \psi(\xi_j) + a - \sum_{i=1}^r w_i \phi(\xi_j|\theta^{(i)}) \right]^2$$

$$= \sum_{j=1}^q (\psi(\xi_j) + a)^2 + w^T \left( \sum_{j=1}^q \phi(\xi_j|\Theta)\phi^T(\xi_j|\Theta) \right) w$$

$$- 2w^T \sum_{j=1}^q (\psi(\xi_j) + a)\phi(\xi_j|\Theta),$$

subject to  $\int_0^1 u(\theta)d\theta = 1$ .

The problem of minimization of the function  $f_a$  is equivalent to minimization of the following function

$$\tilde{f}_a = \frac{1}{2} w^T A w - b^T w + c,$$

where

$$A = \sum_{j=1}^q \phi(\xi_j|\Theta)\phi^T(\xi_j|\Theta), \quad b = \sum_{j=1}^q (\psi(\xi_j) + a)\phi(\xi_j|\Theta),$$

$$c = \frac{1}{2} \sum_{j=1}^q (\psi(\xi_j) + a)^2.$$

To minimize with respect to  $a$  we use the MATLAB function `fminbnd` within the interval  $[-10, 10]$ . To minimize  $f_a$  for each fixed  $a$  we shall use some recent gradient methods described in the next section.

### 3.3 Gradient algorithms

Consider a problem of minimizing a quadratic function in the following form

$$f(x) = \frac{1}{2} x^T A x - x^T b, \tag{4}$$

where  $x \in \mathbb{R}^r$  is an unknown vector,  $b$  is a given vector in  $\mathbb{R}^r$  and  $A$  is a  $r \times r$  symmetric positive-definite matrix such that

$$0 < m = \inf_{z^T z=1} z^T A z < M = \sup_{z^T z=1} z^T A z < \infty,$$

The gradient of the function (4) at point  $x_k$  is  $\nabla f(x_k) = g_k = Ax_k - b$ . The solution to the optimization problem (4) is  $x^* = A^{-1}b$ . However, usually it is a very difficult computational task to compute the inverse matrix  $A^{-1}$ , especially when the dimension of the problem,  $r$ , is large. In this case some numerical methods can be used to determine the solution  $x^*$ . A big class of numerical optimization methods is the class of gradient algorithms.

The general gradient method for solving problem (4) corresponds to the following iterative process. Let  $x_0 \in \mathbb{R}^r$  be a starting vector and  $g_0 = Ax_0 - b \in \mathbb{R}^r$  be the initial residual vector. At step  $k = 1, 2, \dots$ , let

$$x_{k+1} = x_k - \alpha_k g_k, \tag{5}$$

where  $\alpha_k > 0$ , the step-size at iteration  $k$ , is determined by some rule. The iterations (5) can also be rewritten in terms of residuals (gradients)  $g_k$  as

$$g_{k+1} = g_k - \alpha_k A g_k.$$

There exist numerous algorithms of gradient-type, the most popular is the Steepest Descent method (also known as

Gradient Descent), nowadays regarded as an algorithm with poor rate of convergence. The Barzilai and Borwein method introduced by Barzilai and Borwein (1988) has much better convergence rate, but it is completely non-monotonic. For more details on the construction of sequence of step lengths of this type and the behavior of associated gradient methods see Zhigljavsky et al. (2013).

### 4 Illustrative examples

In this Section we present several numerical examples to illustrate the approach. Using iterative gradient methods, we are trying to minimize objective function (3) for  $\psi(\xi) = -IMSPE(x|\xi)$  and

$$\phi_{1*}(\xi|\theta) = |M_1(\theta)|^{\theta_3} |M_2(\theta)|^{1-\theta_3}, \quad 0 \leq \theta_3 \leq 1.$$

For this criterion the resulting matrix  $A$  is usually very ill-conditioned. This can lead to a slow convergence of gradient methods. The speed of convergence depends strongly on the spectrum of  $A$ , especially on the condition number (ratio of the smallest and the largest eigenvalues of the matrix). In this case the smallest eigenvalue is very close to zero and the largest one is usually of magnitude  $10^5$ . It is known that when the condition number is large, numerical methods converge slowly.

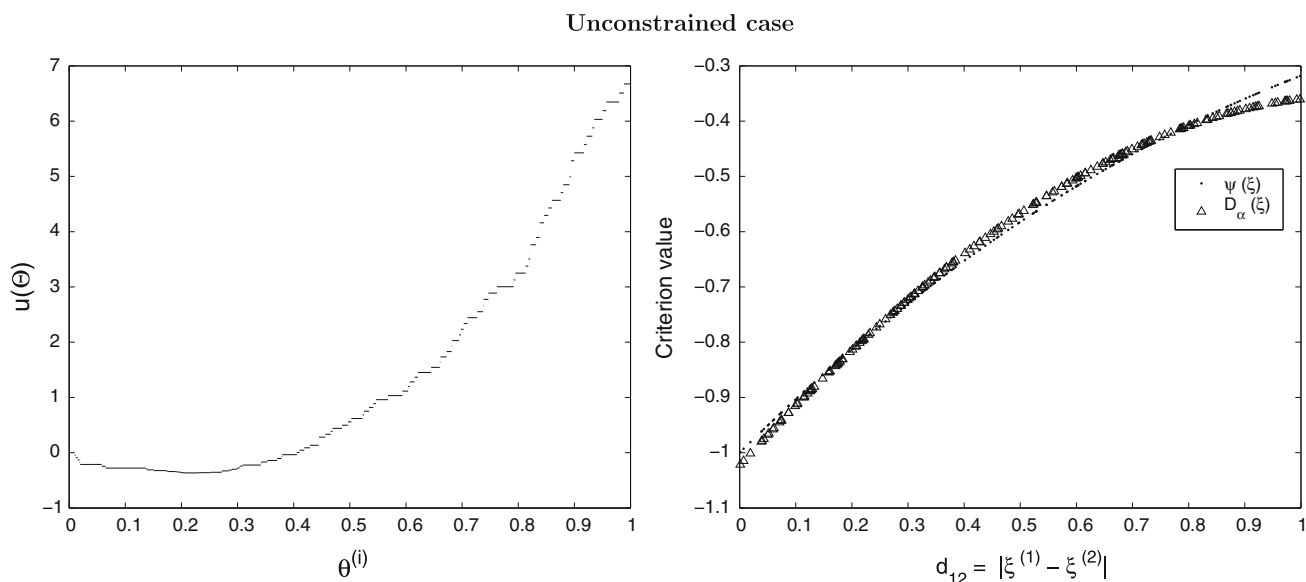
When using other criterion function

$$\phi_2(\xi|\theta) = \theta_3 \log |M_1(\theta)| + (1 - \theta_3) \log |M_2(\theta)|, \quad 0 \leq \theta_3 \leq 1,$$

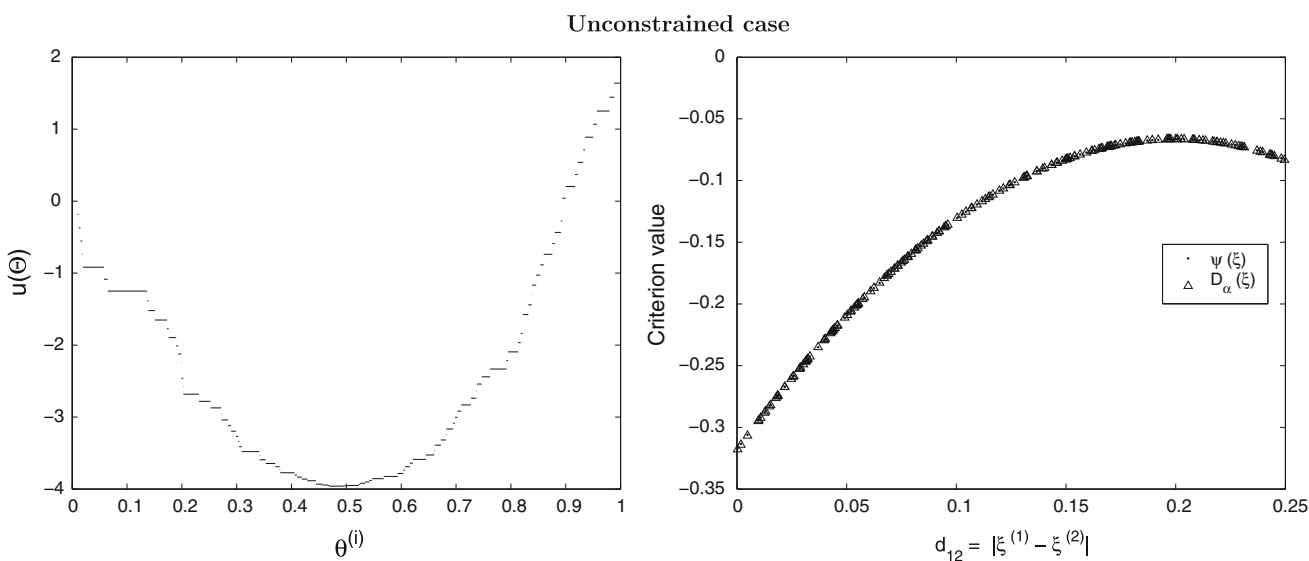
the condition number of  $A$  is still quite large, however, it is smaller than in the previous case. Thus, for our computations we have used a log-based criterion function.

For the examples below we choose the dimension of the partition  $\Theta$ , which is also the dimension of the problem, to be  $r = 100$ . Thus, matrix  $A$  is a symmetric  $r \times r$  matrix with the largest eigenvalue and the smallest eigenvalue very close to zero. For the gradient algorithm the initial vector is  $w^{(0)} = (1, 1, \dots, 1)$  and the stopping criterion is  $\|g\|_2 \leq tol$  with  $tol = 10^{-4}$ . For each example we plot the resulting function given by (2), and the two criteria as a function of  $d_{12} = |\xi^{(1)} - \xi^{(2)}|$ , where  $\xi^{(1)}$  and  $\xi^{(2)}$  are the first and second components of the corresponding design.

We have used 2, 3, 4, 5 and 6-point designs  $\xi$  of the following type:  $\xi_j = (0, x_2^{(j)})$ ,  $\xi_j = (0, x_2^{(j)}, 1)$ ,  $\xi_j = (0, \frac{1}{2}x_2^{(j)}, x_2^{(j)}, 1)$ ,  $\xi_j = (0, \frac{1}{3}x_2^{(j)}, \frac{2}{3}x_2^{(j)}, x_2^{(j)}, 1)$ ,  $\xi_j = (0, \frac{1}{4}x_2^{(j)}, \frac{1}{2}x_2^{(j)}, \frac{3}{4}x_2^{(j)}, x_2^{(j)}, 1)$  with  $1 \leq j \leq 200$ , where  $x_2^{(j)}$  is a random vector. In each example the parameter of the covariance of the process  $\theta_2$  is equal to one. For the sake of brevity we plot only results for 2 and 6-point designs.



**Fig. 2** Values of  $w_i, i = 1, \dots, r$ , against  $\Theta = (\theta^{(1)}, \dots, \theta^{(r)})$  (left) and  $\psi(\xi)$  and  $D_\alpha(\xi)$  as a function of  $d_{12} = |\xi^{(1)} - \xi^{(2)}|$  (right) for 2-point designs.



**Fig. 3** Values of  $w_i, i = 1, \dots, r$ , against  $\Theta = (\theta^{(1)}, \dots, \theta^{(r)})$  (left) and  $\psi(\xi)$  and  $D_\alpha(\xi)$  as a function of  $d_{12} = |\xi^{(1)} - \xi^{(2)}|$  (right) for 6-point designs.

4.1 Unconstrained case

First, we do not impose any additional constraints on  $u(\theta)$ . For the illustration see Figs. 2 and 3. We plot the resulting values of  $w_i, i = 1, \dots, r$ , against  $\Theta$ , and the two criteria  $\psi(\xi)$  and  $D_\alpha(\xi)$  versus the distances of the first two points of the designs. In the figures  $D_\alpha$  denotes  $\int_0^1 \phi(\xi|\theta)u(\theta) d\theta$ , where as  $u(\theta)$  we use the optimal function that we have found. In this case, approximations are successful: in each

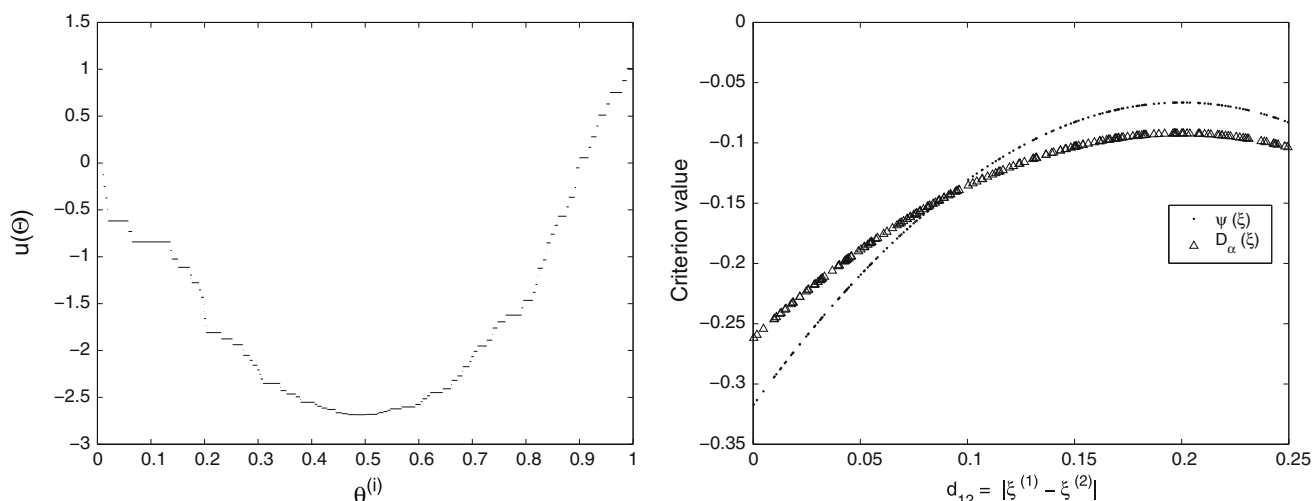
example  $D_\alpha(\xi)$  is very close to  $\psi(\xi)$ , even for the case of 2-point designs.

4.2 Constrained case

Now we consider that the function  $u(\theta)$  is constrained so that

$$\int_0^1 u(\theta)d\theta = 1,$$

Constrained case



**Fig. 4** Values of  $w_i, i = 1, \dots, r$ , against  $\Theta = (\theta^{(1)}, \dots, \theta^{(r)})$  (left) and  $\psi(\xi)$  and  $D_\alpha(\xi)$  as a function of  $d_{12} = |\xi^{(1)} - \xi^{(2)}|$  (right) for 6-point designs.

or simply  $\sum_{i=1}^r w_i = 1$ . In this case we apply the gradient method to minimize the function  $f_\sigma$  described in the previous Section (see Sect. 3.1 for more details), with the penalty term  $\sigma = 10$ . See Fig. 4 for illustration.

In the constrained case, we were not able to construct reasonable approximations of  $\psi(\xi)$ . Perhaps, this happened because of the constraints imposed on  $u(\theta)$ . For the optimal  $u(\theta)$  which satisfies the constraints and minimizes the function  $f_\sigma$ , the relative distance (the value of  $f_\sigma$  at the optimum  $u(\theta)$ ) is not small enough to provide a good fitting of the two criteria. This problem can be fixed by introducing a scaling factor,  $a$ , as described in Sect. 3.2. This approach is implemented in the upcoming sections.

### 4.3 Weighted and unconstrained case

In this section we present the results of the minimization of a function  $f_a$  with a scaling factor  $a$ , see Sect. 3.2 for the explanation. First, we shall illustrate the weighted and unconstrained case. It corresponds to the case described in the Sect. 3.2 when the constraints on  $u(\theta)$  are omitted. See Figs. 5 and 6 for the results.

As in the unconstrained case we were able to obtain reasonable approximations of  $\psi(\xi)$ . Moreover, in this case minimization was also done with respect to the scaling factor  $a$ , thus we have slightly better results than in the simple case. This is due to the more complex computations as well as an additional minimization with respect to the weight  $a$ . Of course, this minimization requires some extra calculations. For each fixed  $a$  the corresponding unconstrained problem is solved. From this point of view, the first example in the Sect. 4.1 can be regarded as a unconstrained problem that corresponds to the scaling factor  $a = 0$ . Then,

the weight  $a$ , for which  $f_a$  is reaching the minimum value, is chosen. Thus, minimization with respect to  $a$  consists in a sequential solution of unconstrained problems and then choosing the best of them. As a result, the computational time is increased but the precision of approximation is better.

It can be seen from the figures that for each optimal scaling factor (Table 1), the two criteria  $\psi(\xi)$  and  $D_\alpha(\xi)$  are very close to each other.

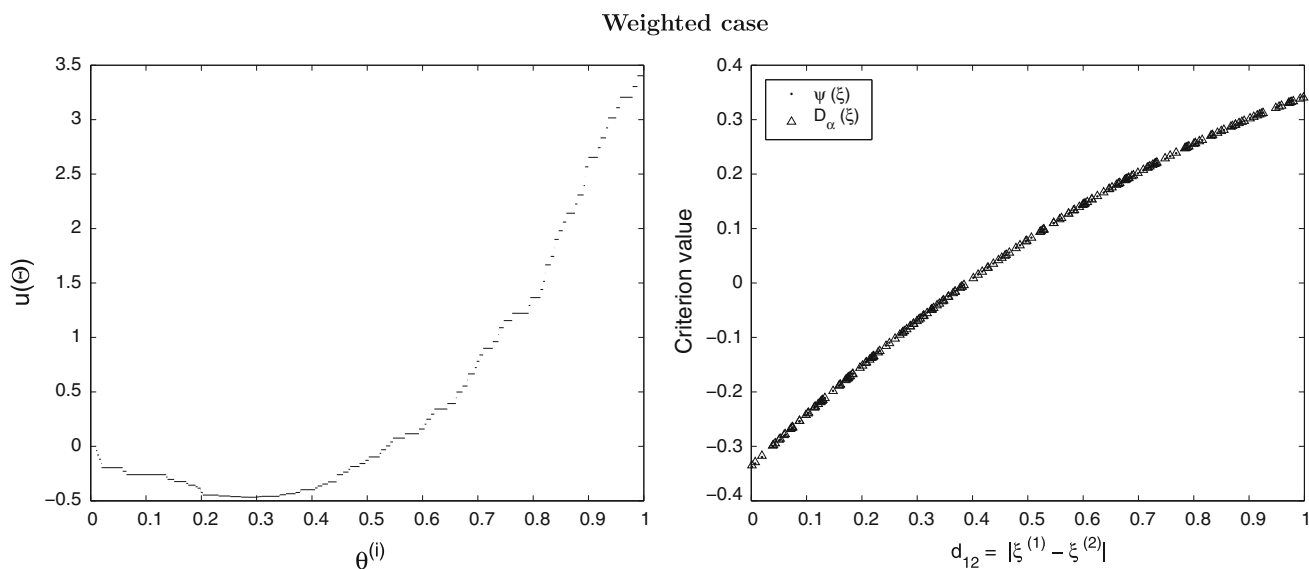
### 4.4 Weighted and constrained case

Let us now illustrate the weighted and constrained case, see Figs. 7 and 8. Unlike in the simple constrained case, we were able to construct reasonable approximations when some weight was added to the objective function. Thus, our experiments showed that scaling factor  $a = 0$  is not always optimal. Table 1 summarizes the optimal values of scaling factors for the different cases that we have considered. Thus, introduction of the weight allowed us to approximate  $\psi(\xi)$  even in the case when some constraints on  $u(\theta)$  were imposed.

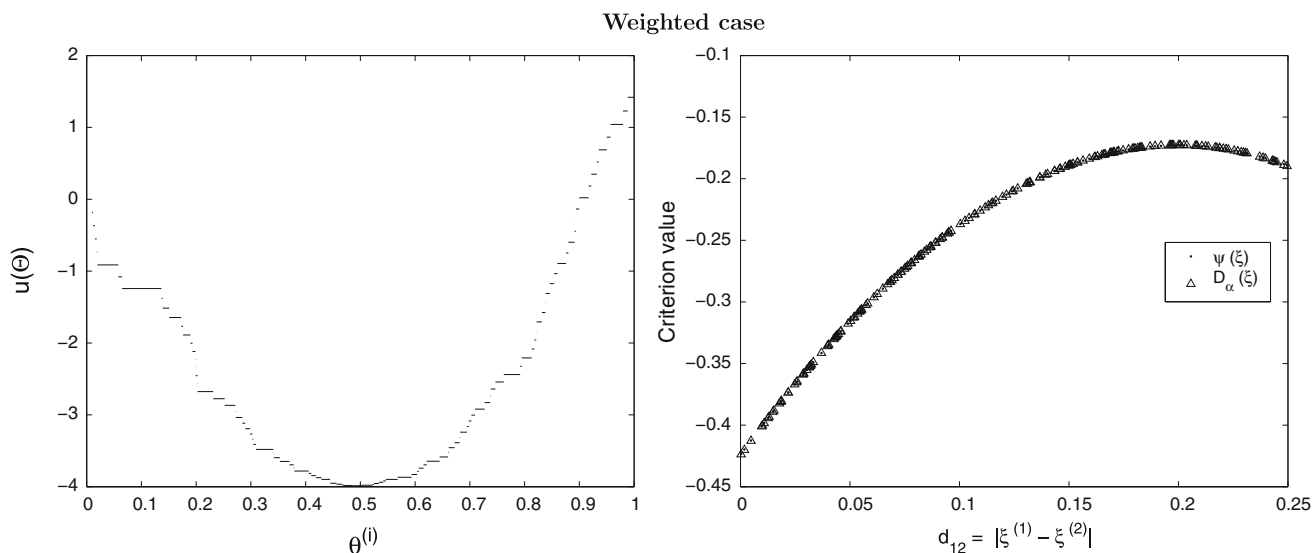
The value of the minimized function at the optimum can be treated as a relative distance between the two criteria, see Table 2. Smaller this distance is better the approximation.

### 4.5 Degenerated case

In all the previous examples the partition  $\Theta = (\theta^{(1)}, \dots, \theta^{(r)})$  was taking as a random vector. However, even if we take a partition  $\Theta$  in a very simple form, say  $\Theta = (0, \dots, 0, 1)$ , we still can obtain reasonable approximation of the criterion  $\psi(\xi) = -IMSPE(x|\xi)$ . One of the



**Fig. 5** Values of  $w_i, i = 1, \dots, r$ , against  $\Theta = (\theta^{(1)}, \dots, \theta^{(r)})$  (left) and  $\psi(\xi)$  and  $D_\alpha(\xi)$  as a function of  $d_{12} = |\xi^{(1)} - \xi^{(2)}|$  (right) for 2-point designs.



**Fig. 6** Values of  $w_i, i = 1, \dots, r$ , against  $\Theta = (\theta^{(1)}, \dots, \theta^{(r)})$  (left) and  $\psi(\xi)$  and  $D_\alpha(\xi)$  as a function of  $d_{12} = |\xi^{(1)} - \xi^{(2)}|$  (right) for 6-point designs.

possible explanations for this phenomenon is the relative simplicity of the chosen criterion *IMSPE*. Perhaps, when using a more sophisticated criterion, say *MSPE*, a simple  $\theta$  will not be reliable. This aspect is left for further investigation.

**5 Simultaneous equations and conditionally restricted designs**

In this section an example with two different approaches to a problem with an explanatory variable,  $z$ , that is not under

the control of the experimenter and it can only be observed after the experiment is realized, are considered. There is another variable,  $x$ , that has to be designed. The conditionally restricted (CR) approach [López-Fidalgo and Garcet-Rodríguez (2004), Martín-Martín et al. (2007)] considers a joint design of the two variables,  $(x, z)$  taking into account that the second can not be designed and is going to be observed after the experiment. The restriction is in fixing a conditional design for the first variable given the second, say  $\xi_{z|x}(\cdot|x)$ . That is, what is expected to happen with the first variable if the experiment is this or that. This design is assumed completely known and has to be

**Table 1** Values of the optimal scaling factor  $a$

Designs $\xi$	Unconstrained	Constrained
2-Point designs	0.6612	1.1438
3-Point designs	1.6488	1.4385
4-Point designs	1.5914	0.9933
5-Point designs	0.8209	0.3862
6-Point designs	-0.1064	-0.3097

**Table 2** Distances between the two criteria

Designs $\xi$	No constraints	Weight	Constraints and weight
2-Point designs	$3.7700 \times 10^{-2}$	$1.5450 \times 10^{-4}$	$2.03300 \times 10^{-2}$
3-Point designs	$1.2204 \times 10^{-5}$	$1.2750 \times 10^{-8}$	$2.1008 \times 10^{-7}$
4-Point designs	$1.3000 \times 10^{-3}$	$1.3660 \times 10^{-6}$	$1.8261 \times 10^{-4}$
5-Point designs	$2.0676 \times 10^{-4}$	$4.7275 \times 10^{-6}$	$6.1369 \times 10^{-5}$
6-Point designs	$6.8409 \times 10^{-6}$	$3.5032 \times 10^{-6}$	$1.5682 \times 10^{-5}$

guessed from the experience. Under this approach an information matrix,  $M_{CR}$ , is obtained. From the simultaneous equations models (SE) point of view [see e.g. Conlisk (1979)] the uncontrolled variable is modeled through a new probability distribution, which has to be fitted as well. In this approach one explanatory variable (exogenous) of the main equation,  $z$ , is then the response variable (endogenous) of a second equation. In both equations there is a controllable variable,  $x$ . Thus, another FIM,  $M_{SE}$ , is obtained. Casero-Alonso and López-Fidalgo (2014) provided details about all this. In this case the correlation structure is not defined for the different observations but for the two potential response variables in the model.

In particular, the response  $y$  is modeled through a probability distribution with a pdf depending on a variable  $z$  (only known once the experiment is realized) and the designable variable  $x$ ,  $g(y|x, z; \theta_1)$ , with mean  $\eta_1(x, z; \theta_1)$  and variance  $\sigma_y^2$ . Now two ways of dealing with  $z$  are

1. *CR* There is prior information about the unknown values of  $z$  given  $x$  through a completely known conditional probability design,  $\tilde{\xi}_{z|x}(\cdot|x)$ .
2. *SE* A second model for  $z$  through a marginal pdf  $h(z|x; \theta_2)$ . In this case there are two sets of parameters,  $\theta = (\theta_1^T, \theta_2^T)^T$ .

Let the design space be  $\chi = \{0, 1\}$ , let  $\mathcal{Z} = \{0, 1\}$  be the space of possible outcomes of  $z$ , and assume a normal distribution for the response with

$$E(y|x, z) = \eta_1(y|x, z; \theta_1) = a + bx + cz$$

for both approaches. The unknown variable  $z$  is designed (CR) / modeled (SE) by the probability of  $z = 1$  given a value  $x$ , say  $r_x$ . In this example the value of the probability of  $z = 1$  is assumed known for  $x = 0$ ,  $r_0 = 0$ . This means no action and so no reaction at all. Meanwhile the probability of  $z = 1$  for  $x = 1$ ,  $r_1$ , is to be estimated through the experimentation. Thus, the vector of parameters is  $\theta^T = (\theta_1^T, \theta_2)$  with  $\theta_1^T = (a, b, c) \in \mathbb{R}^3$  and  $\theta_2 = r_1 \in [0, 1]$ . For simplicity, the variance will be assumed known,  $\sigma_y^2 = 1$ .

A general design for  $\chi$  is a two-point design,

$$\xi = \begin{Bmatrix} 0 & 1 \\ 1-p & p \end{Bmatrix}, 0 \leq p \leq 1.$$

Since the model is not linear the FIM depends on the parameter  $\theta_2 = r_1$ . Computing the FIMs in the usual way

$$M_{SE}(\xi) = \begin{pmatrix} 1 & p & pr_1 & 0 \\ p & p & pr_1 & 0 \\ pr_1 & pr_1 & pr_1 & 0 \\ 0 & 0 & 0 & \frac{p}{r_1(1-r_1)} \end{pmatrix} = \begin{pmatrix} M_{CR}(\xi) & 0 \\ 0 & M_{II}(\xi) \end{pmatrix},$$

where the subscripts of the matrices stand for the two different approaches considered.

Now we use the operators given in the Introduction to compute compound designs following the procedure given in this paper. For  $\phi_1$  the determinants cancel and a criterion defined on the determinant of the second diagonal matrix  $|M_{II}|$  is obtained. The function  $u_{\alpha, \beta}(r_1)$  will be considered as a prior Beta distribution with parameters  $\alpha$  and  $\beta$ . We will consider the following kernel, based on the efficiencies,

$$\phi_3(\theta, M_{CR}(\theta), M_{II}(\theta)) = r_1 \left( \frac{|M_{CR}(\theta)|}{|M_{CR}^*(\theta)|} \right)^{1/3} + (1-r_1) \frac{|M_{II}(\theta)|}{|M_{II}^*(\theta)|},$$

where  $M_{CR}^*(\theta)$  and  $M_{II}^*(\theta)$  are the information matrices maximizing the corresponding determinants. In particular, the maximum of

$$|M_{CR}(\theta)| = (1-p)p^2r_1(1-r_1)$$

is  $|M_{CR}^*(\theta)| = \frac{4}{27}r_1(1-r_1)$  at  $p_{CR}^* = 2/3$  and the maximum of

$$|M_{II}(\theta)| = \frac{p}{r_1(1-r_1)}$$

is  $|M_{II}^*(\theta)| = \frac{1}{r_1(1-r_1)}$  at  $p_{II}^* = 1$ . Then

Weighted and constrained case

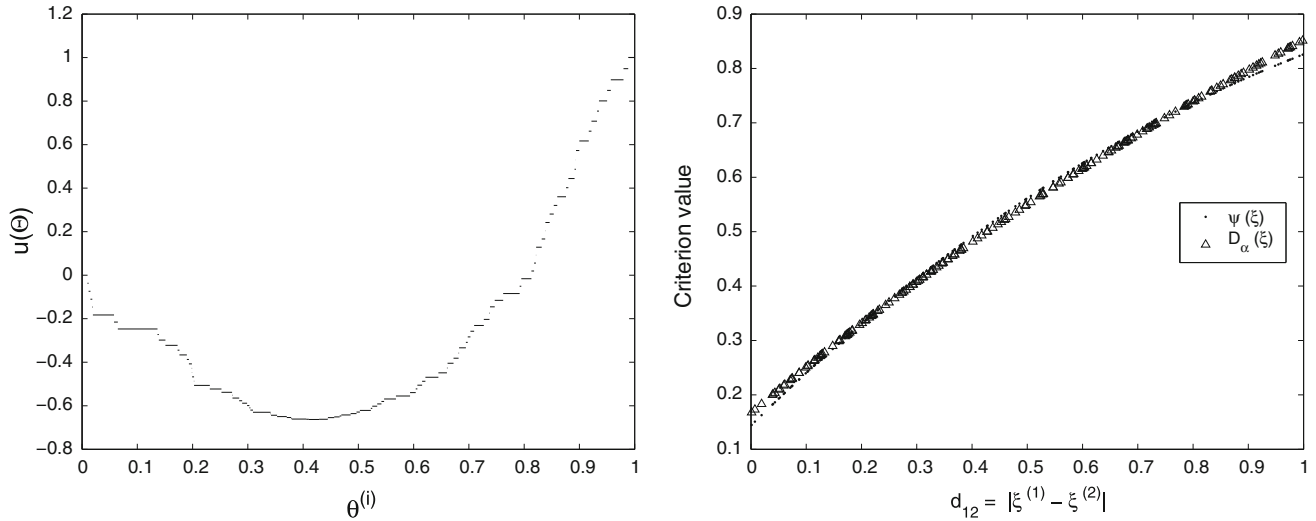


Fig. 7 Values of  $w_i, i = 1, \dots, r$ , against  $\Theta = (\theta^{(1)}, \dots, \theta^{(r)})$  (left) and  $\psi(\xi)$  and  $D_\alpha(\xi)$  as a function of  $d_{12} = |\xi^{(1)} - \xi^{(2)}|$  (right) for 2-point designs.

$$\phi_3(\theta, M_{CR}(\theta), M_{II}(\theta)) = r_1 \left[ \frac{27p^2(1-p)}{4} \right]^{1/3} + (1-r_1)p$$

and

$$Lu(p) = \int_0^1 \phi_3(r_1, M_{CR}(r_1), M_{II}(r_1)) u_{\alpha,\beta}(r_1) dr_1 = \frac{\beta p + \frac{3}{2^{2/3}} \alpha [p^2(1-p)]^{1/3}}{\alpha + \beta}.$$

It will be assumed that the probability of  $z = 1$  (e.g. desaturation in blood) in the experimental group (e.g. doing an exercise,  $x = 1$ ) is 0.4 on average with a variance of 0.04. This means that  $\alpha = 2$  and  $\beta = 3$ . Then the value of  $p$  maximizing  $Lu(p)$  is  $p^* = \frac{1}{3} \left( 2 - \frac{1}{\sqrt[3]{1+\sqrt{2}}} + \sqrt[3]{1+\sqrt{2}} \right) = 0.8654$ . Efficiencies are

$$\text{Eff}_{II}(p_{CR}^*) = \frac{\frac{2/3}{r_1(1-r_1)}}{\frac{1}{r_1(1-r_1)}} = 2/3 = 0.6667$$

$$\text{Eff}_{II}(p^*) = \frac{\frac{0.8654}{r_1(1-r_1)}}{\frac{1}{r_1(1-r_1)}} = 0.8654$$

$$\text{Eff}_{CR}(p_{II}^*) = 0$$

$$\text{Eff}_{CR}(p^*) = \left( \frac{(1-0.8654)0.8654^2 r_1(1-r_1)}{\frac{4}{27} r_1(1-r_1)} \right)^{1/3} = 0.879551$$

The efficiencies of the integrated compound design are quite robust for both matrices.

5.1 A methane case study

Rodríguez-Díaz et al. (2012) considered a case study on methane (CH<sub>4</sub>) removal in the atmosphere, a very important issue at this moment. The reaction of methane with the hydroxyl radical can be expressed in the Modified-Arrhenius (MA) form as

$$k(t) = a \frac{1}{t^m} \exp\left(-\frac{\beta}{t}\right),$$

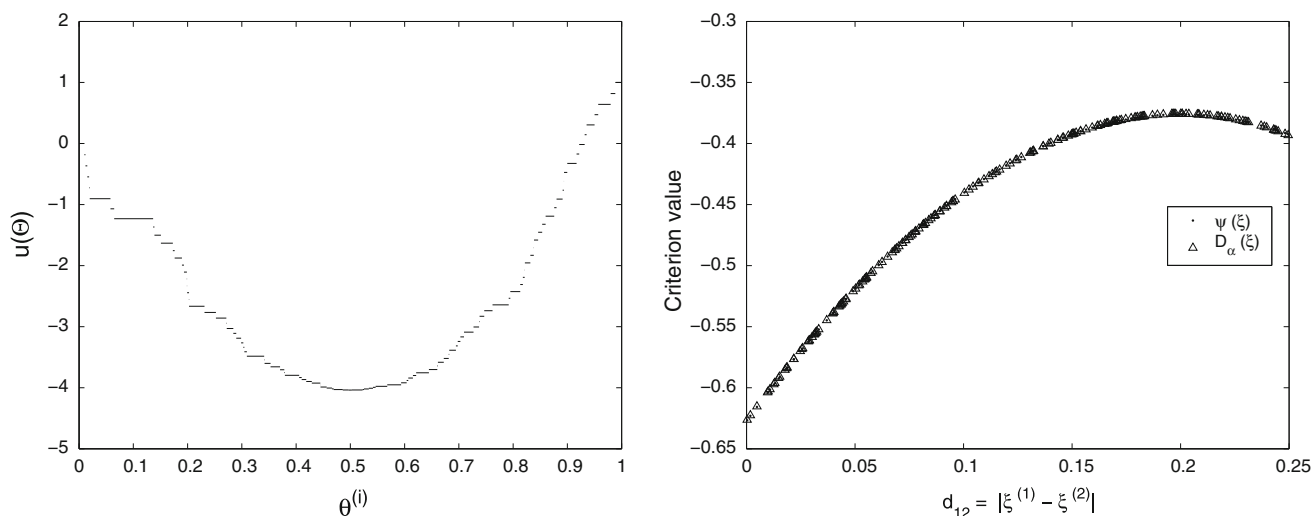
with covariance structure  $\text{cov}(t_i, t_j, r) = \sigma^2 \exp(-dr) = \sigma^2 \exp(-|t_i - t_j|r), r > 0$ . Fitting the parameters of this model resulted in  $a_0 = 2.80 \times 10^{-14}$ ,  $m_0 = -0.667$  and  $\beta_0 = 1575$  for temperatures between 295 and 660 K. For the calculations the Generalized Exponential model, which is equivalent to the MA model through the change of variable  $t_i = 1/x_i$ , is considered. Let us assume  $m$  known and consider two point-designs of the form  $(x, x + d)$ . The FIM for both parameters of interest  $\theta = (\theta_1; r) = (a, \beta; r)$  is

$$M(\theta) = \begin{pmatrix} M_{\theta_1}(\theta) & 0 \\ 0 & M_r(\theta) \end{pmatrix}.$$

As in the previous Sect. 5 the kernel  $\phi_3$  is considered

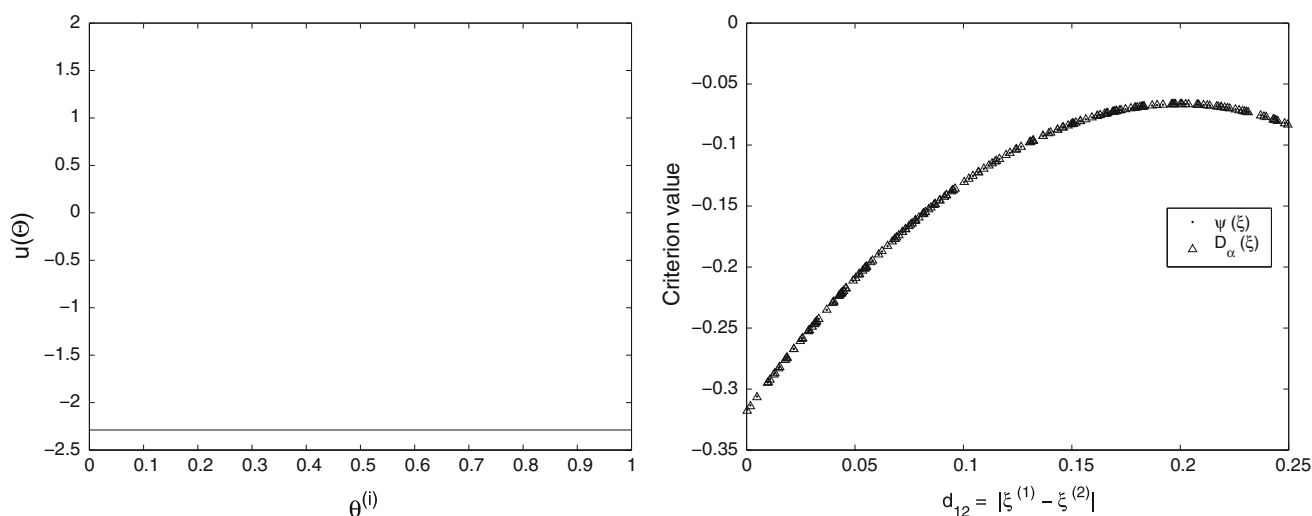
$$\phi_3(\theta, M_{\theta_1}(\theta), M_r(\theta)) = r \left( \frac{|M_{\theta_1}(\theta)|}{|M_{\theta_1}^*(\theta)|} \right)^{1/2} + (1-r) \frac{|M_r(\theta)|}{|M_r^*(\theta)|}.$$

Weighted and constrained case



**Fig. 8** Values of  $w_i, i = 1, \dots, r$ , against  $\Theta = (\theta^{(1)}, \dots, \theta^{(r)})$  (left) and  $\psi(\xi)$  and  $D_\alpha(\xi)$  as a function of  $d_{12} = |\xi^{(1)} - \xi^{(2)}|$  (right) for 6-point designs.

Degenerated case



**Fig. 9** Values of  $w_i, i = 1, \dots, r$ , against  $\Theta = (\theta^{(1)}, \dots, \theta^{(r)})$  (left) and  $\psi(\xi)$  and  $D_\alpha(\xi)$  as a function of  $d_{12} = |\xi^{(1)} - \xi^{(2)}|$  (right) for 6-point designs.

The maximum of

$$|M(\theta_1)| = \frac{a^2 d^2 \exp[2dr - 2\beta(2x + d)]x^{2m}(x + d)^{2m}}{\exp(2dr) - 1}$$

for the nominal values of the parameters is reached at the maximum distance  $d_{\theta_1}^* = 1/660 - 1/295 = 0.00187468$ . On the other hand, the maximum of

$$|M(r)| = \frac{d^2 [1 + \exp(2dr)]}{[\exp(2dr) - 1]^2}$$

is reached at  $d_r^* = 0$ . Then

$$\phi_3(\theta, M_{\theta_1}(\theta), M_r(\theta)) = r \left( \frac{|M_{\theta_1}(\theta)|}{|M_{\theta_1}^*(\theta)|} \right)^{1/2} + (1 - r) \frac{|M_r(\theta)|}{|M_r^*(\theta)|}$$

For the function  $u(r)$  any distribution with  $r > 0$  would be considered. For the calculations the distributions Weibull and Gamma for different values of their parameters are considered. The same results are obtained. The value of  $d$  maximizing

$$Lu(d) = \int_0^1 \phi_3(r, M_{\theta_1}(r), M_r(r)u_{\alpha,\beta}(r))dr,$$

is the maximum distance  $d^* = 0.001875$ . Then the IMSPE optimal design matches the D-optimal design for the  $\theta_1$  parameters. This design has a high efficiency, near 1, with respect to the nominal value of  $r$  since  $d^*$  is rather small.

## 6 Discussion

Integral compounding of two matrices can help to construct designs that are similarly informative for two different models. In Sect. 5 a binary design space is considered and a procedure is proposed in order to find compound designs in an easy and efficient way. The method is based on the minimization of a function  $\phi$ , which measures discrepancy between information matrices of the two models. Different functions are considered. The comparison of rival models does not necessarily mean that they share any of the parameters. One example of this is the Ornstein–Uhlenbeck process for which the trend and the correlation structure have different parameters. Amo-Salas et al. (2013) studied the case when common parameters exists, in particular for covariance structures proportional to the trend. They showed this is not a standard case and they gave conditions and tools to model in practice the covariance in such a way. In the case of two rival models, it might happen that parameters are the same. However we have not studied such a case in the present paper and this is a valuable further research direction. A common weighting is used to regularize the discrepancy through a density defined on parameter  $p$ . Such a compounding in the limit case may cover classical compound information criteria. The process is illustrated with relevant examples. Assuming two different approaches to deal with an explanatory variable that is only known after the experiment is performed. The designs obtained in both examples have high efficiencies. In this paper we have considered the case of simple two point design measure. This can be generalized for arbitrary  $n$ -point designs using a Dirichlet density for the design weights, however further research will be needed to fulfill this aims.

The procedure was applied to a real example modeling a flux of methane in troposphere. The Kyoto Protocol aimed to reduce emissions of methane. The fall in human-induced methane emissions in the 1990s was only transitory Lelieveld (2006). We should be aware of deterministic, stochastic and chaotic parts of physical and biological systems and we need robust designs for measurement of such a complex environmental systems. This paper aims to provide a new way for construction of such designs, motivated

by information aggregation, given by a simple additive Lebesgue integral.

In Sects. 2 and 3 the operator approximation for learning about the relationship between IMSPE and the compound criterion  $D_x$  was given. First we introduced the integral relationship between IMSPE and the compounding kernel. We provide numerical implementation and construct the solution  $u$  of a given Fredholm integral equation of the first kind. Several numerical examples illustrates the methods provided.

In general, it is a difficult task to make a link between two optimality criteria: one based on prediction, another based on parameter estimation, especially in the case of correlated processes. In the uncorrelated case it was done by Kiefer and Wolfowitz (1960). The correlated case, however, is not fully investigated. The first attempt to relate two criteria was done in Müller and Pronzato (2009). In this paper we have concentrated on the correlated case and we have reformulated the conjecture of equivalence in an integral form.

Initially, we reformulated the problem of equivalence as a quadratic optimization problem. Then we solved this problem by using new gradient methods introduced in Zhigljavsky et al. (2013) for quadratic optimization. These methods are advantageous for several reasons: they are much faster than standard gradient methods, have small computational cost, they are easily parallelized and monotonic at almost every iteration. We tried the proposed approach on several test problems and we presented the obtained results in the paper. In terms of required computational effort it is less expensive to generate optimal designs for parameter estimation rather than optimal designs for prediction. Some particular structure of the relationship between two criteria would allow to substitute a costly computation by a much less costly one.

A thorough study of other regularizations will be of interest. As we have seen in Sect. 4 we need regularization of  $Lu = \psi$ , as it is usual for Fredholm equations of the first kind. We have regularized it by  $\|u\| = 1, u > 0$ . This is a natural form of regularization, since we have  $u$  in the form of density. Regularization by maximum entropy (Amato and Hughes (1991)) and regularization by reproducing kernel Hilbert space norm, (Wahba (1977)) would be worth further investigation.

Table 1 shows the optimal values for the scaling parameter. A sensitivity analysis with respect to this parameter would be of much interest, but this is a rather difficult issue. One problem is that a theoretical proof of weak/strong solvability is difficult, since we cope with Fredholm integral equation of the 1st kind with asymmetric kernel. We would prefer to make a thorough theoretical study of the problem before a sensitivity study, since too many imprecisions are involved (e.g. calibration, chaos in

numerics or stochastics of data). In this paper the value of “ $a$ ” is the output of the `fminbnd` function for the particular settings and its calibration is not completely stochastic, since chaotical aspects of optimization are also present.

Criteria for discrimination between rival models, e.g. T- or KL-optimality [Atkinson and Fedorov (1975), Atkinson and Fedorov (1975), López-Fidalgo et al. (2007)] may be included under this approach with some generalizations not considered in this paper. Assuming two rival models with pdf’s  $f_1(y; x, \theta)$  and  $f_2(y; x, \theta)$  the kernel will be now a function which depend on the FIM’s of each model only intrinsically,

$$\phi(\theta, M_1(\theta), M_2(\theta)) = \int f_1(y; x, \theta) \log\left(\frac{f_1(y; x, \theta)}{f_2(y; x, \theta)}\right) dy,$$

and the function  $u$  vanished everywhere except for  $\theta^* = \arg \max_{\theta \in \Theta} \phi(\theta, M_1(\theta), M_2(\theta))$ , where  $u(\theta^*) = 1$ . Here this approach does not add much, but theoretically it may be considered as a particular case.

**Acknowledgments** This work was partially done while V. Casero-Alonso visited the Institute of Statistics of Johannes Kepler University. He wants to thank their hospitality and ideas. This work has been supported by Ministerio de Educación y Ciencia and Fondos FEDER MTM2010-20774-C03-01 and Junta de Comunidades de Castilla la Mancha PEII10-0291-1850 and Amadee, Project Nr. FR 11/2010. The work of E. Bukina was partially supported by the EU through a Marie-Curie Fellowship (EST-SIGNAL program: <http://est-signal.i3s.unice.fr>) under the contract Nb. MEST-CT-2005-021175. Milan Stehlik was supported by ANR project Desire FWF I 833-N18. The authors are thankful for helpful comments of Werner G. Müller, Luc Pronzato and Joao Rendas. We also thank the editor and reviewers, whose insightful comments helped us to sharpen the paper considerably.

**Appendix: Proofs and technicalities**

**Proposition 1** Let us have

$$Y(x_i) = \eta(x_i, \vartheta) + \varepsilon(x_i), \quad i = 1, 2, \quad x_1, x_2 \in [0, 1],$$

$$\gamma(d) = 1 - \exp(-\theta d),$$

where  $\gamma$  stands for the semi-variogram and only the covariance parameter  $\theta$  is of interest. Then the maximal FIM is obtained for  $d = 0$ .

*Proof* We have the log-likelihood function  $L = K - \frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} v^T \Sigma(\theta)^{-1} v$ , where  $v = (Y(x_1) - \eta(x_1, \vartheta), Y(x_2) - \eta(x_2, \vartheta))^T$ . The FIM for the covariance parameter  $\theta$  is

$$M_\theta = E\left(-\frac{\partial^2 L}{\partial \theta^2}\right) = \frac{1}{2} \frac{\partial^2 \log |\Sigma(\theta)|}{\partial \theta^2} + \frac{1}{2} E\left(v^T \frac{\partial^2 \Sigma(\theta)^{-1}}{\partial \theta^2} v\right),$$

and we have  $\frac{\partial^2 \{A_{ij}\}}{\partial \theta^2} \{ \frac{\partial^2 A_{ii}}{\partial \theta^2} \}$  and  $|\Sigma(\theta)| = 1 - \exp(-2\theta d)$ . Further we have

$$\frac{1}{2} \frac{\partial^2 \log |\Sigma(\theta)|}{\partial \theta^2} = -\frac{2d^2 \exp(-2d\theta)}{(1 - \exp(-2d\theta))^2}$$

and

$$\frac{1}{2} E\left(v^T \frac{\partial^2 \Sigma(\theta)^{-1}}{\partial \theta^2} v\right) = \frac{d^2 \exp(-2d\theta)(\exp(-2d\theta) + 3)}{(1 - \exp(-2d\theta))^2}$$

and finally

$$M_\theta = \frac{d^2 \exp(-2\theta d)(1 + \exp(-2\theta d))}{(1 - \exp(-2\theta d))^2}.$$

Note that for every  $\theta > 0$  the maximum  $\frac{1}{2\theta^2}$  is attained for  $d = 0$ .

**Proposition 2** Let us have

$$Y(x_i) = \eta(x_i, \vartheta) + \varepsilon(x_i), \quad i = 1, 2, \quad x_1, x_2 \in [0, 1]$$

$$\text{cov}(x_1, x_2) = 1 - \theta|x_1 - x_2|,$$

and only covariance parameter  $\theta$  is parameter of interest. For regularity assumption we suppose that  $\theta d < 2, \theta > 0$ . Then the maximal FIM is obtained for maximal  $d$ .

*Proof* We have  $|\Sigma(\theta)| = \theta d(2 - \theta d)$ ,

$$\frac{1}{2} \frac{\partial^2 \log |\Sigma(\theta)|}{\partial \theta^2} = -\frac{-2\theta d + \theta^2 d^2 + 2}{\theta^2(\theta d - 2)^2}$$

and

$$\frac{1}{2} E\left(v^T \frac{\partial^2 \Sigma(\theta)^{-1}}{\partial \theta^2} v\right) = 2 \frac{-2\theta d + \theta^2 d^2 + 2}{\theta^2(\theta d - 2)^2}$$

and finally

$$M_\theta = \frac{-2\theta d + \theta^2 d^2 + 2}{\theta^2(\theta d - 2)^2}.$$

We have

$$\frac{\partial M_\theta}{\partial d} = \frac{2d}{(2 - \theta d)^3}$$

So  $M_\theta$  is increasing function for every (acceptable)  $0 < d < \min\{\frac{2}{\theta}, 1\}$ .

**Proposition 3** By direct integration we obtain the following:

$$Lu^* = \int_0^1 \varphi(M_{\text{exp}}(\theta), M_{\text{lin}}(\theta), \theta) u^*(\theta) d\theta$$

$$= A/B,$$

where

$$\begin{aligned}
A &= -6 \exp(2\theta d)\theta^2 d^2 + 4 \exp(2\theta d)\theta d - \theta^4 d^4 \exp(2\theta d) \\
&\quad + 4\theta^3 d^3 \exp(2\theta d) - 2\theta d \exp(4\theta d) + \theta^2 d^2 \exp(4\theta d) \\
&\quad + 2 + 2 \exp(4\theta d) - 4 \exp(2\theta d) - \theta^4 d^4 + 4\theta^3 d^3 \\
&\quad - 3\theta^2 d^2 - 2\theta d, \\
B &= (-\log((-2\theta d + \theta^2 d^2 + 2)/\theta^2/(\theta^2 d^2 - 4\theta d + 4))\theta^2 d^2 \\
&\quad + 4 \log((-2\theta d + \theta^2 d^2 + 2)/\theta^2/(\theta^2 d^2 - 4\theta d + 4))\theta d \\
&\quad - 4 \log((-2\theta d + \theta^2 d^2 + 2)/\theta^2/(\theta^2 d^2 - 4\theta d + 4)) \\
&\quad + \log(-d^2(\exp(2\theta d) + 1)/(-\exp(4\theta d) + 2 \exp(2\theta d) \\
&\quad - 1))\theta^2 d^2 - 4 \log(-d^2(\exp(2\theta d) + 1)/(-\exp(4\theta d) \\
&\quad + 2 \exp(2\theta d) - 1))\theta d + 4 \log(-d^2(\exp(2\theta d) + 1)/ \\
&\quad (-\exp(4\theta d) + 2 \exp(2\theta d) - 1))).
\end{aligned}$$

## References

- Ahmadi J, Arghami NR (2003) Comparing the Fisher information in record values and iid observations. *Stat A J Theor Appl Stat* 37(5):435–441
- Alshunnar FS, Raqab MZ, Kundu D (2012) On the comparison of the Fisher information of the log-normal and generalized Rayleigh distributions. *J Appl Stat* 37(3):391–404
- Amato U, Hughes W (1991) Maximum entropy regularization of Fredholm integral equations of the first kind. *Inverse Probl* 7:793–808
- Amo-Salas M, López-Fidalgo J, Porcu E (2013) Optimal designs for some stochastic processes whose covariance is a function of the mean. *Test* 22:159–181
- Atkinson AC, Fedorov VV (1975) The designs of experiments for discriminating between two rival models. *Biometrika* 62:57–70
- Atkinson AC, Fedorov VV (1975) Optimal design: experiments for discriminating between several models. *Biometrika* 62:289–303
- Baldi Antognini A, Zagoraiou M (2010) Exact optimal designs for computer experiments via Kriging metamodelling. *J Stat Plan Inference* 140:2607–2617
- Barzilai J, Borwein JM (1988) Two-point step size gradient methods. *IMA J Numer Anal* 8:141–148
- Casero-Alonso V, López-Fidalgo J (2014) Experimental designs in triangular simultaneous equations models. *Stat Pap* (in press)
- Conlisk J (1979) Design for simultaneous equations. *J Econom* 11(1):63–76
- Cook RD, Wong WK (1994) On the equivalence of constrained and compound optimal designs. *J Am Stat Assoc* 89(426):687–692
- Crary SB (2002) Design of computer experiments for metamodel generation. *Analog Integr Circuits Signal Process* 32:7–16
- Hansen PC (1992) Numerical tools for analysis and solution of Fredholm integral equations of the first kind. *Inverse Probl* 8:849–872
- Hofmann G (2004) Comparing the Fisher information in record data and random observations. *Stat Pap* 45(4):517–528
- Kiefer J, Wolfowitz J (1960) The equivalence of two extremum problems. *Can J Math* 12:363–366
- Kiseľák J, Stehlík M (2008) Equidistant D-optimal designs for parameters of Ornstein–Uhlenbeck process. *Stat Probab Lett* 78:1388–1396
- Lelieveld J (2006) A nasty surprise in the greenhouse. *Nature* 443:405–406
- López-Fidalgo J, Garcet-Rodríguez S (2004) Optimal experimental designs when some independent variables are not subject to control. *J Am Stat Assoc* 99:1190–1199
- López-Fidalgo J, Tommasi C, Trandafir PC (2007) An optimal experimental design criterion for discriminating between non-normal models. *J R Stat Soc B* 69(2):231–242
- Martín-Martín R, Torsney B, López-Fidalgo J (2007) Construction of marginally and conditionally restricted designs using multiplicative algorithms. *Comput Stat Data Anal* 51:5547–5561
- McGree JM, Eccleston JA, Duffull SB (1988) Compound optimal design criteria for nonlinear models. *J Biopharm Stat* 18(4):646–661
- Müller WG, Pronzato L (2009) Towards an optimal design equivalence theorem for random fields? IFAS report Nr. 45 of the Department for Applied Statistics of the Johannes Kepler University in Linz
- Müller WG, Stehlík M (2009) Issues in the optimal design of computer simulation experiments. *Appl Stoch Models Bus Ind* 25:163–177
- Müller WG, Stehlík M (2010) Compound optimal spatial designs. *Environmetrics* 21:354–364
- Pázman A (2010) Information contained in design points of experiments with correlated observations. *Kybernetika* 46(4):771–783
- Rodríguez-Díaz JM, Santos-Martín MT, Waldl H, Stehlík M (2012) Filling and D-optimal designs for the correlated generalized exponential models. *Chemometr Intell Lab Syst* 114:10–18
- Sacks J, Schiller SB, Welch WJ (1989) Design for computer experiments. *Technometrics* 31(1):41–47
- Tandeo P, Ailliot P, Autret E (2011) Linear Gaussian state-space model with irregular sampling: application to sea surface temperature. *Stoch Environ Res Risk Assess* 25:793–804
- Tikhonov AN, Arsenin VY (1977) Solutions of ill-posed problems. Wiley, New York
- Unami K, Abagale FK, Yangyuoru M, Alam AHMB, Kranjac-Berisavljevic G (2010) A stochastic differential equation model for assessing drought and flood risks. *Stoch Environ Res Risk Assess* 24:725–733
- Wahba G (1977) Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J Numer Anal* 14:651–667
- Zhigljavsky AA, Pronzato L, Bukina E (2013) An asymptotically optimal gradient algorithm for quadratic optimization with low computational cost. *Optim Lett* 7(6):1047–1059

## Optimal designs subject to cost constraints in simultaneous equations models

Víctor Casero–Alonso · Jesús  
López–Fidalgo

the date of receipt and acceptance should be inserted later

**Abstract** A procedure based on a multiplicative algorithm for computing optimal experimental designs subject to cost constraints in simultaneous equations models is presented. A convex criterion function based on a usual criterion function and an appropriate cost function is considered. A specific L–optimal design problem and a numerical example are taken from Conlisk (1979) to compare the procedure. The problem would need integer non linear programming to obtain exact designs. To avoid this he solves a continuous non linear programming problem and then he rounds–off the number of replicates of each experiment. The procedure provided in this paper reduces dramatically the computational efforts computing optimal approximate designs. It is based on a specific formulation of the asymptotic covariance matrix of the full–information maximum likelihood estimators, which simplifies the calculations. The design obtained for estimating the structural parameters of the numerical example by this procedure is not only easier to compute, but also more efficient than the design provided by Conlisk.

**Keywords** Approximate design, Cost constraints, Exact design, L–optimal design, Multiplicative algorithm, Simultaneous equations, Structural equations.

**Mathematics Subject Classification (2000)** 62K05, 62P20

### 1 Introduction

Optimal experimental designs when constraints are imposed have been considered in the literature widely (see e.g. Cook and Fedorov 1995). In this paper the

---

Department of Mathematics, Institute of Mathematics Applied to Science and Engineering, University of Castilla–La Mancha, Avda. Camilo José Cela 3, 13.071 Ciudad Real, Spain. E-mail: Victormanuel.Casero@uclm.es Phone: +34 926 295 300 (ext. 6402). Fax: +34 926 295 361

problem is focused on the total cost of the experiment. A scalar coefficient will tune the compromise between information and cost (benefit). The novelty is to consider this approach for a model of simultaneous equations. These models are very common in the economics literature among other areas.

As far as the authors know the first who considered the problem of designing experiments for simultaneous equations models was Conlisk (1979). For him the natural approach, as in standard regression, is to minimize a scalar function of the covariance matrix of the estimated structural coefficients of the simultaneous equations model. He proposed a design objective function based either on the three–stage least squares (3SLS) estimators introduced by Zellner and Theil (1962), or on the full–information maximum likelihood (FIML) estimators. These are the most commonly used estimation procedures for these models. The FIML method is computationally expensive as many authors pointed out. But the 3SLS method needs the reduced–form of the simultaneous equations model, that is, the set of equations that expresses each endogenous variable solely in terms of the exogenous variables. In the structural form of the model each endogenous variable is specified as a regression over the exogenous and the endogenous variables. Rothenberg and Leenders (1964) gave an explicit expression for asymptotic covariance matrix of FIML estimators which is equivalent to the asymptotic covariance matrix of 3SLS estimators of the structural coefficients, assuming the normal distribution for the disturbances. Finally, Conlisk pointed out that the nonlinear restrictions on the reduced–form coefficients implied by the structural model causes the design criteria to be functions of unknown parameters. For handling these nonlinear restrictions Papakyrizakis (1986) suggested the combination of sequential estimation and design control strategies.

There is no much literature about optimal design for simultaneous equations models. Casero–Alonso and López–Fidalgo (2014) proposed an alternative approach using triangular simultaneous equations models to the optimal experimental design problem for a linear regression model with explanatory variables that are not subject to the control of the practitioner. They obtained optimal designs for this model. Here it is extended to a problem with cost constraints. Casero–Alonso and López–Fidalgo (2014) calculated the Information Matrix as the expected value of the partial derivative of the loglikelihood function with respect to the parameters of a triangular model, FIML method. In this paper the direct expression of the asymptotic covariance matrix of the FIML estimators provided by Rothenberg and Leenders (1964) is used for a more general model and an efficient iterative procedure is proposed to find optimal designs.

In order to compute optimal designs we consider a multiplicative algorithm based on a convex criterion function. The multiplicative algorithm (e.g. Torsney and Martín–Martín 2009) is used to update simultaneously all weights of the design points, step by step. A modification of the weighted average of an optimality criteria and a cost function (Pronzato 2010) is considered in order to fit better the procedure. Cook and Wong (1994) showed that constrained

and compound optimal designs approaches to handling multiple objectives are equivalent.

This approach is based on approximate designs. But an exact design is the final objective. The difference is not trivial. An exact design will be a collection of points  $x_1, \dots, x_n$  for a fixed number of experiments,  $n$ . Some of these points may be repeated. A discrete probability can be defined from it assigning the proportion of repetitions of each point in the design. Extending this, Kiefer (1959) defined an approximate design as a probability measure on  $\chi$ , say  $\xi$ . This definition provides a framework for constructing optimal approximate designs and for discussing various optimality properties of designs in linear models using tools from convex analysis.

The organization of this paper is as follows. In Section 2 the simultaneous equations model and the optimal experimental design problem in that model subject to a cost constraint is defined. In Section 3 a procedure to solve optimal design problems using a multiplicative algorithm, is presented. In Section 4 L-optimal designs subject to cost constraints are obtained in a couple of numerical examples. A discussion concludes this paper.

## 2 Optimal experimental designs in simultaneous equations models

In this section a general problem of designing an optimum experiment for a simultaneous equations model subject to cost constraints is formulated. First the usual notation of simultaneous equations is introduced.

Let us suppose a complete system of  $r$  linear stochastic structural equations in  $r$  jointly dependent variables and  $k$  predetermined explanatory variables,

$$y_i = Y_i\alpha_i + X_i\beta_i + u_i = Z_i\delta_i + u_i, \quad i = 1, \dots, r, \quad (1)$$

where  $y_i$  is the  $n \times 1$  vector of observations of one of the endogenous dependent variables,  $n$  is the number of experiments,  $Y_i$  is the  $n \times r_i$  matrix of values formed by the  $r_i \leq r - 1$  explanatory endogenous variables included in that equation,  $X_i$  is the  $n \times k_i$  matrix of values formed by the  $k_i \leq k$  explanatory exogenous variables included in that equation,  $\alpha_i$  and  $\beta_i$  are  $r_i \times 1$  and  $k_i \times 1$  structural parameters vectors to be estimated, for the included endogenous and exogenous variables respectively, and  $u_i$  is the  $n \times 1$  vector of structural disturbances that have zero mean and are independent and homoscedastic across observations, that is with common covariance matrix,  $\text{cov}(u_{1h}, \dots, u_{rh}) = \Sigma$ ,  $h = 1, \dots, n$ . Let  $y_i = Z_i\delta_i + u_i$  where  $Z_i = (Y_i, X_i)$  and  $\delta_i = (\alpha_i, \beta_i)$ .

Furthermore, let us suppose that the reduced-form of the structural equations system exists,  $Y = X\Pi + V$ , where  $Y$  is the matrix of endogenous variable observations,  $X$  is the matrix of values of the exogenous variables,  $\Pi$  is the coefficients matrix, and  $V$  is the reduced-form disturbances matrix. Let  $Y_i = X\Pi_i + V_i$  where  $\Pi_i$  and  $V_i$  are submatrices of  $\Pi$  and  $V$  for that part of the reduced-form corresponding to  $Y_i$ .

As mentioned above the most common methods to estimate the structural coefficients  $\delta = (\alpha_1, \beta_1, \dots, \alpha_r, \beta_r)^T$  of (1) are 3SLS, three-stage least squares, and FIML, full-information maximum likelihood. The 3SLS method (Zellner and Theil 1962) estimates the coefficients in three stages. The first serves to estimate the moment matrix of the reduced-form disturbances. In the second stage the coefficients of one single structural equation are estimated. Using the two-stage least squares estimated moment matrix of the structural disturbances, all coefficients of the entire system are estimated simultaneously in the third stage.

The asymptotic covariance matrix of 3SLS or FIML estimators of structural coefficients  $\delta$  of (1) assuming the normal distribution for the disturbances is

$$\text{plim}(\text{cov } \hat{\delta}) = \left[ -n \lim_{n \rightarrow \infty} E \left( \frac{\partial^2 L^*}{\partial \delta \partial \delta^T} \right) \right]^{-1} = \{H^T [\Sigma^{-1} \otimes (X^T X)] H\}^{-1}, \quad (2)$$

where plim indicates the probability limit,  $L^*$  is the ‘concentrated’ likelihood function (see Rothenberg and Leenders 1964),  $\otimes$  represents the Kronecker product of matrices and

$$H = \begin{pmatrix} H_1 & 0 \\ & \ddots \\ 0 & H_r \end{pmatrix}, \quad (3)$$

where  $H_i = (II_i, J_i)$  and  $J_i$  is defined by  $X_i = XJ_i$ , the matrix of unit vectors that select from  $X$  the exogenous variables of the  $i$ th equation.

## 2.1 A design problem

Once the model is defined a set of specifications to state the optimal experimental design problem is needed: the design space  $\mathcal{X}$ , the experimenter’s objective function and the cost constraints, if any. Although the design space could be continuous, the search is usually restricted to finite designs (Karlin and Studden 1966). Regarding the objective function, any of the classical criteria of optimal experimental design, as D-optimality, A-optimality, c-optimality..., could be considered.

To illustrate the procedure introduced in this paper, a specific L-optimal design problem is considered. A L-optimal design  $\xi^*$  is a design that minimizes the trace of  $LM(\xi)^{-1}$  (Atwood 1976), where  $M(\xi)$  is the Information Matrix associated to an approximate design  $\xi$  and  $L$  is a definite positive user-selected matrix.  $L$  is chosen to reflect the researcher’s interest in the study. In the example considered here, the experimenter’s objective function is the sum of estimation error variances for predicting  $y_i$ ,  $i = 1, \dots, r$ , that is, the predicted values of the left-hand variable of each structural equation, evaluated at each of the points of the finite design space  $\mathcal{X} = \{x_1, \dots, x_m\}$ . The cost constraint is  $\sum_{j=1}^m c_j n_j \leq C$ , where  $c_j$  is the cost of taking one observation at  $x_j$ ,  $n_j$  is

the number of times  $x_j$  appears in the final design and  $C$  is the total budget of the experiment.

Putting all this together, the specific L-optimal experimental design problem can be stated as

$$\begin{aligned} \min_{n_1, \dots, n_m} \operatorname{tr} & \left\{ \left[ H^T (I_r \otimes \sum_j x_j^T x_j) H \right] \left[ H^T \left( \Sigma^{-1} \otimes \sum_j n_j x_j^T x_j \right) H \right]^{-1} \right\} \\ \text{s.t.} & \sum_j n_j c_j \leq C \text{ and } n_1 \geq 0, \dots, n_m \geq 0, \end{aligned} \quad (4)$$

where  $I_r$  is the identity matrix of order  $r$ . As Conlisk (1979) pointed out, though the  $n_j$  must be integers (exact design), the usual procedure is to approximate by solving the continuous nonlinear programming problem and rounding-off the optimal values of  $n_j$ .

### 3 An efficient iterative procedure

In this section a procedure to solve the experimental design problem in a simultaneous equations model subject to a cost constraint stated in (4) is established. It is an alternative to the computationally expensive continuous nonlinear programming problem.

Given that  $X^T X$  of (2) can be expressed as  $\sum_j n_j x_j^T x_j$  in (4), that is, as a weighted sum of the information matrices of the support points, approximate designs may be used here. Then a multiplicative algorithm for approximate designs can be applied. At each step the weights of the design points,  $\xi(x_j)$ , will be adjusted. Therefore, one of the advantages of the procedure described below is the reduction in time of computation (see a comparison of computational times in Section 5).

Problem (4) is equivalent to obtaining the optimal approximate design  $\xi^*$  that minimizes the convex criterion function (Cook and Wong 1994)

$$\Phi_\beta(\xi) = \Phi(\xi) - \beta B(\xi), \quad (5)$$

where  $\xi$  is an approximate design,  $\Phi$  is an optimality criterion function (usually a function of  $M^{-1}(\xi)$ ),  $B(x_j)$  is an appropriate benefit function based on the known cost function  $c_j$ ,  $B(\xi) = \sum_i \xi(x_j) B(x_j)$  which is linear in  $\xi$  and  $\beta > 0$  is a parameter to be tuned. Note that the sign minus previous to  $\beta$  is to obtain the maximum benefit/minimum cost.

The General Equivalence Theorem, GET (Kiefer and Wolfowitz 1959) says that  $\xi^*$  is  $\Phi_\beta$ -optimal if and only if

$$\partial \Phi_\beta(\xi, 1_{x_j}) = \operatorname{tr} \left\{ \nabla \Phi(\xi) [M(1_{x_j}) - M(\xi)] \right\} - \beta [B(1_{x_j}) - B(\xi)] \geq 0, \quad j = 1, \dots, m,$$

where  $\partial\Phi(\xi, \xi')$  is the directional derivative of  $\Phi$  at  $\xi$  in the direction of  $\xi'$ , that is

$$\partial\Phi(\xi, \xi') = \lim_{\varepsilon \rightarrow 0^+} \frac{\Phi[(1 - \varepsilon)\xi + \varepsilon\xi'] - \Phi(\xi)}{\varepsilon}$$

and  $1_{x_j}$  is the design concentrated at  $x_j$ , i.e. with weight 1 at  $x_j$ .

Based on the GET a multiplicative algorithm is used (Torsney and Mart n-Mart n 2009). For a specific value of  $\beta$ , at step  $s + 1$  the weights of the design are updated,

$$\xi^{(s+1)}(x_j) = \xi^{(s)}(x_j) \frac{\beta B(1_{x_j}) - \text{tr} [\nabla\Phi(\xi^{(s)})M(1_{x_j})]}{\beta B(\xi^{(s)}) - \text{tr} [\nabla\Phi(\xi^{(s)})M(\xi^{(s)})]}. \quad (6)$$

With this formula  $\sum_j \xi^{(s+1)}(x_j) = 1$ .

For setting a stopping rule, a lower bound of the efficiency of the new design  $\xi^{(s+1)}$  is computed. Again it is based on the GET, and it depends on the sign of  $\Phi_\beta(\xi^{(s+1)})$ ,

- If  $\Phi_\beta(\xi^{(s+1)}) > 0$  then  $\text{eff}_{\Phi_\beta} [M(\xi^{(s+1)})] \geq 1 + K$ ,
- If  $\Phi_\beta(\xi^{(s+1)}) < 0$  then  $\text{eff}_{\Phi_\beta} [M(\xi^{(s+1)})] \geq 1/(1 + K)$ ,

where  $K = \min_{x_j} \partial\Phi_\beta(\xi^{(s+1)}, 1_{x_j})/\Phi_\beta(\xi^{(s+1)})$ .

For computing the optimal design  $\xi^*$  for the criterion (5) the following procedure is established:

1. For each particular value of  $\beta$  an optimal approximate design is obtained numerically,

$$\xi^{(\beta)} = \arg \min_{\xi} \Phi_\beta(\xi) = \arg \min_{\xi} [\Phi(\xi) - \beta B(\xi)],$$

using the multiplicative algorithm (6).

2. Let  $\beta^* = \arg \min_{\beta} \Phi(n^{(\beta)}\xi^{(\beta)})$ , where  $n^{(\beta)} = C/C(\xi^{(\beta)})$  is the total number of experiments for each  $\beta$ , and  $C(\xi^{(\beta)}) = \sum_j \xi^{(\beta)}(x_j)c_j$ .
3. Finally  $n^{(\beta^*)}\xi^{(\beta^*)}(x_j)$  has to be rounded-off to integers,  $n_j^*$ , in such a way  $\sum_j n_j^* \approx n^{(\beta^*)}$  and  $\sum n_j^*c_j \approx C$ .

This procedure could be adapted to any design problem.

#### 4 An illustrative example

To compare the procedure described above with the continuous problem stated in (4) the numerical example of Conlisk (1979) is considered. First we take a simplified version to illustrate both approaches. Later the full version of the example is commented.

Let us suppose that a major oil company decides to experiment with controlled variation in the prices of gas and repairs,  $P_g$  and  $P_r$ , using its service station outlets as a cross-section of observations. Three alternate values

(0.93, 1, 1.07) for each price variable (in an appropriate normalized form) are considered. This makes a full factorial design space of  $3^2$  experimental conditions:

$$\mathcal{X} = \{x_1 = (0.93, 0.93), x_2 = (0.93, 1), x_3 = (0.93, 1.07), \\ x_4 = (1, 0.93), x_5 = (1, 1), x_6 = (1, 1.07), \\ x_7 = (1.07, 0.93), x_8 = (1.07, 1), x_9 = (1.07, 1.07)\}.$$

The goal is to estimate the coefficients of the demand model (structural equations) of quantities of gas and repairs sold,  $Q_g$  and  $Q_r$ ,

$$\begin{cases} Q_g = \alpha_{11} + \beta_{11}Q_r + \alpha_{12}P_g + u_1, \\ Q_r = \alpha_{21} + \beta_{22}Q_g + \alpha_{22}P_r + u_2. \end{cases}$$

To solve the corresponding experimental design problem (4) we must supply nominal values for  $\Sigma$ , disturbance covariance matrix, and  $H$ , where  $H$  depends on  $\Pi$  via (3) and  $\Pi$  in turn depends on the structural coefficients  $\alpha_i$  and  $\beta_i$ . This is because the nonlinear restrictions on the reduced-form coefficients  $\Pi$  implied by the structural model. The prior specifications needed for computing the optimal design are

$$\begin{cases} Q_g = 1.38 + 0.1Q_r - 0.5P_g + u_1, \\ Q_r = 0.39 + 0.8Q_g - 0.2P_r + u_2, \end{cases} \quad \Sigma = \begin{pmatrix} 0.1 & 0.06 \\ 0.06 & 0.1 \end{pmatrix}.$$

With these values we obtain the reduced-form system after some algebra, and therefore the nominal values of  $\Pi_1$  and  $\Pi_2$ . Then the  $H_i$  matrices are

$$H_1 = \begin{pmatrix} 1.6239 & 1 & 0 \\ -0.4348 & 0 & 1 \\ -0.2174 & 0 & 0 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 1.5424 & 1 & 0 \\ -0.5438 & 0 & 0 \\ -0.0217 & 0 & 1 \end{pmatrix}.$$

Finally, the cost of making one experiment at the  $j$ th experimental condition (service station) is defined by

$$c_j = 5 + 3673(P_g^{(j)} - 1)^2 + 2857(P_r^{(j)} - 1)^2 + 1633(P_g^{(j)} - 1)(P_r^{(j)} - 1),$$

in terms of percentages of profit lost by stations during the experiment (Conlisk 1979). The nine rounded-off unit costs are  $c_j = \{45, 23, 29, 19, 5, 19, 29, 23, 45\}$ . There is symmetry in the unit costs. The minimum unit cost value is in the ‘central’ experimental condition, that is at standard price levels  $P_g = P_r = 1$ . The maximum cost values are in the ‘initial’ and ‘final’ experimental conditions ( $P_g = P_r = 0.93$  and  $P_g = P_r = 1.07$ , respectively). That is, deviations of standard price levels are cost-penalized (yield percent profit losses). The total budget of the experiment is specified as  $C = 6500$ , which is the lost of all profits of 65 stations.

With these specifications we will find numerically the optimal structural design, that is the solution of the design problem (4). In other words, the optimal approximate design for estimating the structural parameters of the simultaneous equations model subject to a cost constraint. There is not a unique global minimum, it depends on the starting point of the search. To illustrate the situation we consider three cases:

	$P_g$			1			1.07		
	0.93	1	1.07	0.93	1	1.07	0.93	1	1.07
Case 1	9	45	19	55	77	55	19	45	9
Case 2	13	37	23	47	92	47	23	37	13
Case 3	19	25	29	36	116	36	29	26	19
P	37	0	39	0	180	0	39	0	37

**Table 1** Optimal exact designs for estimating the structural parameters of the first numerical example. Case 1, 2 and 3 are direct solutions of problem (4). P indicates the solution obtained with the procedure established in this paper.

- Case 1. Initial values of  $n_j = 1$ ,  $j = 1, \dots, m$ .
- Case 2. Initial values of  $n_j = n_j^{\max}/2$ , where  $n_j^{\max} = \arg \max \{n_j | n_j c_j = 6500\}$ ,  $j = 1, \dots, m$ .
- Case 3. Initial values of  $n_j = n_j^{\max} - 1$ .

The optimal exact (rounded-off) designs obtained are presented in Table 1. Observe that when the initial values increase the  $n_j$  of the odd design points decrease whereas the  $n_j$  of the even points increase. Moreover, the ‘central’ design point receives the highest number of experiments. The total number of experiments varies from 332 in Case 2 to 335 in Case 3. For Case 1 is 333. For the three cases the value of the objective function (4) for the approximate design is 0.020973. If the optimal exact designs of Table 1 are considered the values of the objective function for Cases 1, 2 and 3 are 0.0211127, 0.021129 and 0.0209315 respectively. The later value is lower than the value for the optimal approximate design.

The full version of the example considers other controlled variables, the dummies  $D_g$  and  $D_r$ , describing whether or not trading stamps are offered with gas and repair sales respectively. Then, the demand model is:

$$\begin{cases} Q_g = \alpha_{11} + \beta_{11}Q_r + \alpha_{12}P_g + \alpha_{13}D_g + u_1, \\ Q_r = \alpha_{21} + \beta_{22}Q_g + \alpha_{22}P_r + \alpha_{23}D_r + u_2. \end{cases}$$

In this case, the design space  $\mathcal{X}$  is 4-dimensional,  $(P_g, P_r, D_g, D_r)$  with values 0 or 1 for  $D_g$  and  $D_r$ . This yields  $3^2 \times 2^2 = 36$  experimental possible conditions:

$$\mathcal{X} = \{x_1 = (0.93, 0.93, 0, 0), x_2 = (0.93, 0.93, 0, 1), x_3 = (0.93, 0.93, 1, 0), x_4 = (0.93, 0.93, 1, 1), x_5 = (0.93, 1, 0, 0), \dots, x_{36} = (1.07, 1.07, 1, 1)\}.$$

The prior specifications of  $\alpha_{13}$  and  $\alpha_{23}$  are 0.04 and 0.02 respectively. And the cost of the  $j$ th experimental condition, service station, includes now the term  $5(1 - D_g^{(j)})$  to penalize the absence of trading stamps of gas. There is no term to penalize presence/absence of trading stamps of repairs. Therefore each couple of experimental conditions  $(P_g, P_r, D_g, 0)$  and  $(P_g, P_r, D_g, 1)$  have the same unit cost. The rounded-off unit costs are shown in Table 2.

As in the simplified case, there is symmetry in the unit costs. The ‘central’ experimental conditions are the cheapest and the ‘extremes’ ( $P_g$  and  $P_r \neq 1$ ) the most expensive ones.

		$P_r$ 0.93		1		1.07	
		$D_r$ 0 1		0 1		0 1	
$P_g$	$D_g$						
0.93	0	50	50	28	28	34	34
	1	45	45	23	23	29	29
1	0	24	24	10	10	24	24
	1	19	19	5	5	19	19
1.07	0	34	34	28	28	50	50
	1	29	29	23	23	45	45

**Table 2** Rounded-off unit costs

The optimal structural design for this case, provided by Conlisk (1979), is presented in Table 3. This exact design sums up to 529 experiments with a total cost of 6471. We were not able to compute this design directly. But, we have obtained approximate designs similar to the Conlisk design but with a lower value of the objective function (4). This means that the rounding-off procedure is an important issue (see Section 5).

Now the optimal design with the procedure described in Section 3 for the numerical example is obtained. In order to apply it, a criterion function  $\Phi(\xi)$ , the benefit function  $B(\xi)$  and the values of the parameter  $\beta$  have to be selected.

As described in Section 2.1 the criterion function considered is L-optimality:  $\Phi_L[M(\xi)] = \text{tr}[LM^{-1}(\xi)]$ . Then  $\nabla\Phi_L[M(\xi)] = -M^{-1}(\xi)LM^{-1}(\xi)$  for non-singular information matrices. From (2) the FIM is

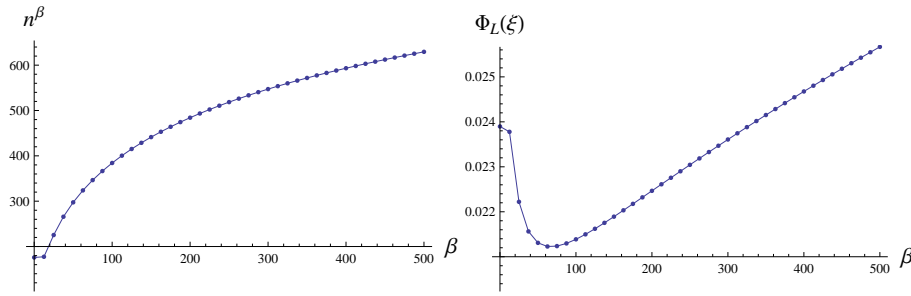
$$M(\xi) = H^T \left( \Sigma^{-1} \otimes \sum_j \xi_j x_j^T x_j \right) H$$

with  $\xi_j = \xi(x_j)$  (such that  $\sum_j \xi_j = 1$ ). From (4)  $L = H^T(I \otimes \sum_j x_j^T x_j)H$ , which is a positive definite matrix of the same dimension as  $M(\xi)$ .

Let the benefit function be  $B(\xi) = \sum_j \xi(x_j)B(x_j)$  where  $B(x_j) = 1/c_j$ . The parameter  $\beta$  depends on the problem considered. For the simplified example the values of  $\beta$  are taken from the interval  $[0, 500]$ .

		$P_r$ 0.93		1		1.07									
		$D_r$ 0 1		0 1		0 1									
$P_g$	$D_g$							$P_g$	$D_g$						
0.93	0	3	1	17	7	6	11	0.93	4	16	0	0	17	6	
	1	1	0	2	0	1	4		1	0	0	0	0	0	0
1	0	17	2	49	57	1	6	1	0	0	0	0	0	0	
	1	8	0	137	136	1	18		1	0	0	165	166	0	0
1.07	0	3	1	0	1	0	0	1.07	0	10	27	0	0	26	9
	1	11	5	5	15	0	3		1	0	0	0	0	0	0
Conlisk (1979)						Procedure of Section 3									

**Table 3** Optimal designs for estimating the structural parameters



**Fig. 1** Values of total number of experiments  $n^{(\beta)}$  and criterion function  $\Phi_L(\xi)$  for different values of  $\beta$  parameter.

Then, the steps of the procedure are applied as described in Section 3. First, for each  $\beta$  the approximate design reaching an efficiency of 99.9% is considered as optimal. Expression (6) for the multiplicative algorithm in this case is

$$\xi^{(s+1)}(x_j) = \xi^{(s)}(x_j) \frac{\beta B(x_j) - \text{tr}[-M^{-1}(\xi^{(s)})LM^{-1}(\xi^{(s)})M(1_{x_j})]}{\beta \sum_j \xi^{(s)}(x_j)B(x_j) - \text{tr}[-M^{-1}(\xi^{(s)})L]}.$$

Among all  $\beta$  we choose  $\beta^* = \arg \min_{\beta} \Phi_L [M(n^{(\beta)}\xi^{(\beta)})]$ . Finally the optimal exact design is obtained by rounding-off the optimal approximate design of  $n^{(\beta^*)}$  experiments.

In the simplified example  $\beta^* = 66.875$  (see Figure 1, right) and the optimal exact design is presented as **P** in Table 1. For the approximate design of size  $n^{(\beta^*)} = 332.08$ ,  $\Phi_L [M(n^{(\beta^*)}\xi^{(\beta^*)})] = 0.021223$ , and for the exact design 0.021222 (slightly better).

For the full example the values of  $\beta$  are taken from the interval  $[0, 2000]$ . Then  $\beta^* = 586.5$  and the optimal exact design is presented as **P** in Table 3. This design sums up to 446 experiments, 16% less than Conlisk exact design, which sums up to 529 experiments but the cost is about the same. In this example,  $\Phi_L [M(n^{(\beta^*)}\xi^{(\beta^*)})] = 0.0907064$  for the approximate design and 0.091391 for the exact design.

Tables 1 and 3 show big differences between the designs obtained through the two approaches. In the designs obtained with the procedure of Section 3 there are many 0's, 4 of 9 in the simplified example, specifically at the even support points, and 26 of 36 in the full example. That is, for both examples the designs are concentrated at the 'extremes' and the central support points, as usual in standard regression. Although the different values, both approaches allocate the highest values of  $n_j$  in the central design points, the cheapest ones.

## 5 Discussion

There are two main advantages of the procedure proposed in this paper. On the one hand, it saves time of computations. The design problem (4) and the alternative procedure have been implemented in *Mathematica*. The time for obtaining any of the Cases 1, 2 or 3 are around 60 seconds, whereas our procedure needs only about 6 seconds. This point is crucial when the number of points increases. As a matter of fact the example of Conlisk (1979) considered 36 experimental conditions (points) (see Section 4).

On the other hand, the designs obtained with our procedure have much larger efficiencies. To compare the designs we compute the relative L–efficiency of a design  $\xi$  with respect to another design  $\xi'$  of reference:

$$\text{eff}_L(\xi, \xi') = \frac{\Phi_L [M(\xi')]}{\Phi_L [M(\xi)]},$$

where exact designs with integers  $n_j$ ,  $j = 1, \dots, m$  are considered. Due to the consideration of these exact designs the magnitudes of information matrices vary depending on the sample size. The relative L–efficiency of the approximate design of 9 support points obtained with the procedure with respect to the Cases 1, 2 or 3 are 98,8%. If exact designs are considered the efficiency drops to 98,6% in the worst case. On the other hand, the exact design of 36 support points obtained with the alternative procedure is 26.6% more efficient than the structural design provided by Conlisk. Note that both designs are L–optimal cost restricted. Thus the efficiency has been computed for the same total cost and not for the same number of experiments as is usually computed in the unrestricted case.

The advantages just mentioned are based on two points. On the one hand, the use of an efficient multiplicative algorithm. On the other hand a convenient definition of the covariance matrix of the estimators of the structural parameters of the model. Moreover, the procedure could be adapted to any design problem.

Another issue is the consideration of approximate designs when the problem considered requires an exact design. The Conlisk approach and our procedure solve the problem using approximate designs, which must be rounded–off. In our procedure the usual rounding–off of the values obtained is taken. But there are many rounding–off methods to be applied to an optimal approximate design (see Imhof et al. 2001 for a list). We have applied the rounding–off methods included in the software BAZI to the approximate solutions obtained. There are nine apportionment methods implemented in BAZI (Maier and Pukelsheim 2007). From divisor methods, with rounding down (Jefferson/D’Hont/Hagenback–Bischoff), rounding up (Adams) or standard rounding (Webster/Sainte–Laguë/Schepers) to quota methods: Hare–quota method with residual fit by greatest remainders (Hamilton/Hare/Niemeyer) and Droop–quota method with residual fit by greatest remainders. The rounded–off designs obtained with BAZI could be compared, if they are different. Therefore, the last step will be to choose the

best one with respect to the criteria considered. For example, for the structural design of 36 support points the Hare-quota method provides an allocation that attains a lower  $\Phi_L(\xi)$  than the standard rounded-off solution of the procedure. The last one is another solution provided by BAZI, the apportionment obtained with the divisor method with standard rounding. The differences between the Hare-quota design and the rounded-off design are in two design points. In  $P_g = P_r = D_g = D_r = 1$ , where Hare-quota method allocates one experiment less than the rounding-off procedure, it goes from 166 to 165 experiments. This experiment ‘shifts’ to the design point  $P_g = P_r = 1.07; D_g = D_r = 0$  where Hare-quota method allocates 27 experiments while the standard rounding-off procedure assigns 26 experiments.

**Acknowledgements** This work has been supported by Ministerio de Educaci3n y Ciencia and Fondos FEDER MTM2010-20774-C03-01 and Junta de Comunidades de Castilla la Mancha PEII10-0291-1850.

## References

- Atwood CL (1976). Convergent design sequences, for sufficiently regular optimality criteria. *Ann Stat.* 4(6):1124–1138.
- BAZI. Berechnung von Anzahlen mit Zuteilungsmethoden im Internet / Calculation of Allocations by Apportionment Methods in the Internet.  
<http://www.uni-augsburg.de/bazi/>
- Casero-Alonso V, L3pez-Fidalgo J (2014). Experimental Designs in Triangular Simultaneous Equations Models. *Stat Pap.* (In press) doi: 10.1007/s00362-014-0581-y
- Conlisk J (1979). Design for simultaneous equations. *J Econometrics.* 11(1):63–76.
- Cook D, Fedorov V (1995). Constrained optimization of experimental design. *Statistics.* 26(2):129–178.
- Cook RD, Wong WK (1994). On the equivalence of constrained and compound optimal designs. *J Am Stat Assoc.* 89(426):687–692.
- Imhof L, L3pez-Fidalgo J, Wong WK (2001). Efficiencies of rounded optimal approximate designs for small samples. *Stat Neerl.* 55:301–315.
- Karlin S, Studden W (1966). Optimal experimental designs. *Ann Math Stat.* 37:783–810.
- Kiefer J (1959). Optimum experimental designs. *J Roy Stat Soc B.* 21:272–319.
- Kiefer J, Wolfowitz J (1959). Optimum design in regression problems. *Ann Math Stat.* 30:271–294.
- Maier S, Pukelsheim F (2007). Bazi A free computer program for proportional representation apportionment. *Preprint Nr. 42/2007. Institut f3r Mathematik, Universit3t Augsburg.*  
<http://opus.bibliothek.uni-augsburg.de/volltexte/2007/711/>
- Papakyriazis PA (1986). Adaptive optimal estimation control strategies for systems of simultaneous equations. *Math Modelling.* 7:241–257.
- Pronzato L (2010). Penalized optimal designs for dose-finding. *J Stat Plan Infer.* 140:283–296.
- Rothenberg TJ, Leenders CT (1964). Efficient estimation of simultaneous equations systems. *Econometrica.* 32:57–76.
- Torsney B, Mart3n-Mart3n R (2009). Multiplicative algorithms for computing optimum designs. *J Stat Plan Infer.* 139:3947–3961.
- Zellner A, Theil H (1962). Three-Stage least squares: simultaneous estimation of simultaneous equations. *Econometrica.* 30:54–78.