Estimación de la severidad de una cartera de préstamos hipotecarios

Gregorio Moral Turiel y Raúl García Baena

1. INTRODUCCIÓN

En los próximos años, la utilización de estimaciones internas de determinados parámetros de las carteras de operaciones de las entidades de crédito va a tener una gran importancia a efectos de asignación de capital económico, cálculo de capital regulatorio y constitución de provisiones. En España, con la entrada en vigor de la provisión estadística y la posibilidad de estimar dichas provisiones mediante «*…métodos de cálculo basados en su propia experiencia de impagos y en las expectativas de pérdidas por categorías homogéneas de riesgo crediticio, teniendo en cuenta… las garantías constituidas y su valor recuperable…*» (1), el reconocimiento de modelos internos para el cálculo de la provisión estadística es ya una realidad normativa.

El desarrollo de los modelos internos, la adaptación de las bases de datos que los nutren, el diseño de sus algoritmos de cálculo y su implementación son uno de los mayores desafíos para las entidades de gran dimensión, que deben aspirar a explotar en su totalidad las posibilidades que el nuevo marco regulatorio y los avances técnicos (tanto en metodologías de medición de riesgo como los puramente tecnológicos) posibilitan, para cumplir con las obligaciones impuestas por los reguladores y para su propia gestión de riesgos.

Todo esto va a suponer también un reto para los supervisores, que van a necesitar referencias y procedimientos de validación de sistemas, de los métodos de cálculo y de los parámetros internos utilizados por las entidades.

El primer gran objetivo en todos estos modelos es determinar la pérdida media anualizada para las principales carteras (grandes empresas, pymes, consumo, hipotecarios con particulares…). Como la media es aditiva, para encontrar dicha pérdida media en una cartera es suficiente ser capaz de calcular la pérdida esperada de cada operación. Dada una definición de incumplimiento, el enfoque más simple consiste en deter-

NOTA: Este artículo es responsabilidad de los autores, quienes agradecen a B. Orsikowsky el interés mostrado en la publicación de este artículo, a M. Pellicer el esfuerzo realizado para hacerlo más legible y a M. Oroz los comentarios realizados.

(1) Punto 8, norma undécima, Circular 4/1991, modificada por las circulares 9/1999 y 4/2000. El subrayado es de los autores.

minar la probabilidad (PD) de que cada operación (agrupadas por subcarteras) incumpla durante un período de un año. Denominando Exp al importe expuesto en el momento del incumplimiento y LGD a la pérdida porcentual de la operación en caso de incumplimiento (2), el producto Exp*LGD representa la pérdida en el instante de incumplimiento. La pérdida (L) en ese período asociada a la operación es cero si no se produce incumplimiento y Exp*LGD en caso contrario. Suponiendo que no existe correlación entre la exposición (Exp) y la LGD, la pérdida esperada (3) es: E(L)=E(PD*Exp*LGD)= PD*E[Exp]*E(LGD).

El factor E(LGD), LGD media de las operaciones (4), aparece de una u otra forma en todos los modelos de estimación de la pérdida. Es, por tanto, fundamental disponer de estimaciones de E(LGD) para carteras adecuadamente segmentadas.

Se han efectuado muchos estudios que tienen como objetivo estimar dicho parámetro para determinadas carteras, principalmente deuda emitida por grandes corporaciones en forma de bonos. Las razones son obvias. Por una parte, el interés de los compradores de bonos en conocer el riesgo y la necesidad de las agencias de calificación de cuantificarlo; por otra, la disponibilidad de datos fiables sobre los incumplimientos y las recuperaciones.

Sin embargo, la situación de otras carteras —como, por ejemplo, la cartera de operaciones hipotecarias con particulares— es totalmente diferente. En este tipo de operaciones es difícil obtener datos fiables sobre los incumplimientos y, más aún, sobre las recuperaciones. Además, debido a particularidades de cada mercado y sistema legal, es muy difícil comparar y/o utilizar resultados obtenidos en países diferentes.

Este trabajo pretende establecer una metodología práctica y rigurosa para estimar la (las) E(LGD) de una cartera de préstamos bancarios hipotecarios a particulares.

Carteras de préstamos con, por ejemplo, gran variabilidad en el LTV (5) de sus operaciones o vida media elevada precisan de una adecuada segmentación para calcular estimaciones de LGD estables en el tiempo. En efecto, si se considera el ejemplo de una cartera segmentada en función del LTV actualizado (6) y se dispone de estimaciones estables en el tiempo para cada uno de los segmentos, se podrán utilizar dichas estimaciones aun cuando cambie la composición de la cartera radicalmente.

---

(2) LGD, del inglés *loss given default;* se utiliza menos el acrónimo LIED, *lost in the event of default.*

(3) Por simplicidad, se incluye el factor de descuento implícitamente en el término E(Exp). En el anejo 1 se detallan explícitamente la influencia del factor de descuento y el tipo de interés de la operación.

(4) También suele utilizarse el término severidad.

(5) LTV *(Loan To Value):* cociente entre el nominal concedido y el valor de tasación del bien en el momento de la concesión.

(6) Definido como el cociente entre el importe actual del préstamo y su valor de tasación.

La definición de incumplimiento (7) utilizada incide directamente en la correspondiente E(LGD). Se ha utilizado una definición basada en el impago durante 90 días, fundamentalmente porque facilita la comparación con otras carteras y/o entidades, y además simplifica la conciliación con los datos contables. Se analizan diferentes definiciones de incumplimiento y la relación entre las E(LGD) asociadas.

Otro factor clave es la definición de pérdida. En este caso, no se ha usado una definición contable, ya que supondría tanto ignorar los costes financieros implícitos como determinados gastos que, contablemente, no se imputan como pérdidas de las operaciones, pero que, económicamente, se deben imputar a dichas operaciones. La LGD asociada a la operación está basada en descontar al momento de incumplimiento las recuperaciones y gastos futuros, y va referida al importe de la deuda en dicho momento. En el caso de recuperaciones por la vía de la adjudicación del bien hipotecado, se ha incorporado una estimación de los gastos necesarios (judiciales y, en su caso, fiscales), ya que en la mayoría de los casos no se dispone del detalle de gastos reales. En los bancos (8) grandes, es frecuente que la adjudicación se haga a una sociedad vinculada. Dadas las limitaciones de la información disponible, a efectos de los cálculos del ejemplo expuesto se ha considerado que el importe por el que se adjudica dicha sociedad el bien, minorado por un determinado coeficiente reductor, es una cantidad recuperada.

Por último, se aplica la metodología desarrollada al caso en España, de una cartera hipotecaria, joven (9) y diversificada, tanto por garantías como geográficamente. A partir de una muestra pequeña, se obtiene una estimación puntual, basada en la media muestral, para E(LGD) del 12,65 %. Se consideran otros posibles estimadores, se estudia la sensibilidad de las estimaciones ante variaciones de los parámetros internos del modelo y se analiza la estabilidad del método de estimación propuesto ante variaciones en la muestra. Asimismo, se obtiene un intervalo de confianza para E(LGD) al 90 % mediante un método no paramétrico. Finalmente, se comparan los resultados obtenidos con referencias *(benchmarks)* y otros estudios sobre LGD.

2. DEFINICIONES DE INCUMPLIMIENTO Y DE PÉRDIDA

2.1. Definición de incumplimiento

La definición de impago (en adelante, incumplimiento) utilizada normalmente por las entidades en sus cálculos internos difiere sustancialmente de la morosidad contable. La razón principal es que las entidades

(7) Se utiliza el término incumplimiento (como traducción del término inglés *default),* frente a otros (impago, mora, fallido), para evitar confusiones con el significado que estos otros términos tienen en los contextos contable, financiero o económico.

(8) Por comodidad, se utiliza la palabra bancos para referirse a todo tipo de entidades de crédito.

(9) La metodología es igualmente aplicable para carteras que no tengan estas características, pero entonces puede ser necesario segmentar la cartera y obtener estimaciones de diferentes LGD.

suelen eliminar de su cómputo de incumplimiento aquellas operaciones en las que el prestatario paga en efectivo la deuda y aquellas procedentes del efecto arrastre (10). También se eliminan normalmente los denominados impagos técnicos, los impagos regularizados posteriormente sin cancelación de la operación y la mayor parte de las reinstrumentaciones. La definición de incumplimiento utilizada incide directamente en la correspondiente LGD. En este trabajo, se ha utilizado una definición basada en el impago durante 90 días, para presentar los resultados finales. Las principales razones para utilizar esta definición de incumplimiento son:

- Facilita la comparación con otras carteras y/o entidades (11).

- Permite comparaciones homogéneas en distintos momentos del tiempo, al tratarse de una definición objetiva.

- Simplifica la comparación (12) con los datos contables.

- Los 90 días de impago implican la clasificación como incumplimiento, según la opción adoptada por España dentro del ejercicio QIS 3 (13).

La variada gama de posibles definiciones de incumplimiento se pueden clasificar en definiciones objetivas y subjetivas.

Por objetivas se entiende aquellas en las que la entrada en incumplimiento depende solo de características observables ajenas a decisiones de la entidad. Las definiciones objetivas pueden parametrizarse, de forma que pueden ser utilizadas durante períodos largos de tiempo o para carteras diferentes, variando los valores de los parámetros. Por ejemplo, una definición basada en el impago de una cantidad superior a un umbral fijo (I), durante un determinado número de días de impago (N), está parametrizada con (I,N) y es fácil entender que tanto el valor de I como el de N pueden ser diferentes para distintos productos. La definición de incumplimiento utilizada (en adelante, DF_1) en la parte práctica de este trabajo (14) es de este tipo con N=90 días y un umbral I para el importe mínimo relevante.

Frecuentemente, las entidades utilizan definiciones subjetivas, que, por ejemplo, dependan de alguna clasificación interna de las operaciones en dificultades, basada en juicios de los gestores de los riesgos o en decisiones que toman las propias entidades (por ejemplo, iniciar un procedimiento judicial).

---

(10) Véase norma 10 de la Circular 4/1991 del Banco de España.

(11) Comparación que está afectada por las diferentes políticas de recobro aplicadas: dos entidades con similares pérdidas, con la misma definición de incumplimiento y distinta intensidad de recobro pueden tener severidades muy diferentes.

(12) Comparación que no es inmediata.

(13) *Paragraph 399, Footnote 74, Quantitative Impact Study Technical Guidance (October 2002)*.

(14) Al tratarse de hipotecarios a personas físicas, las suspensiones de pagos y las quiebras no se incluyen en la definición.

El problema de elegir la mejor definición de incumplimiento para una cartera dada no está resuelto. Generalmente, se tiende a buscar alguna definición de incumplimiento que elimine la mayor parte de las operaciones que finalmente no producen pérdidas. Con frecuencia, estas definiciones van cambiando con el tiempo y tardan en reflejar el deterioro de las operaciones. Estos cambios y la tardanza en reflejar qué operaciones finalmente producirán pérdidas dificultan el cálculo de una LGD útil para estimar las pérdidas de una cartera crediticia, e imposibilitan la comparación homogénea con otras carteras.

2.2. Definición de LGD

Dada una definición de incumplimiento, se define la LGD asociada a una operación clasificada normal (15) como la pérdida porcentual (16) asociada en caso de incumplimiento. La LGD así definida es una variable aleatoria, que, en lo que sigue, se supone sin correlación con la exposición de las operaciones. En este trabajo se está interesado principalmente en la esperanza matemática de esta variable. Para evitar confusiones donde pueda haberlas, se usará explícitamente la notación E(LGD) para referirse a la media de la LGD.

Para estimar la E(LGD) se van a utilizar las pérdidas observadas y/o estimadas de una muestra de operaciones similares (17). Para determinar la muestra base (18), se han obtenido las operaciones que estaban marcadas como DF_1, y de ellas se han extraído las que no habían regularizado su situación mediante pagos en efectivo con anterioridad a esta fecha. Esto es, se ha obtenido una muestra de operaciones que han entrado en incumplimiento bajo otra definición más restrictiva de incumplimiento (en adelante, DF_2). Por lo tanto, lo que se obtiene a partir de la muestra son observaciones de LGD asociadas a DF_2 (en adelante, LGD_2).

La relación entre diferentes definiciones de incumplimiento y sus correspondientes LGD, en un momento dado, puede analizarse mediante la probabilidad condicionada. El gráfico 1, en el que se representan los posibles estados al final del período de vida de una operación que figura como normal en el momento inicial (no está marcada como DF_1), ayuda a entender la relación anterior.

Dadas las dos definiciones de incumplimiento DF_1, DF_2, supóngase que la primera es más amplia que la segunda, es decir, que en cualquier momento, t, entre el conjunto [denotado por $ODF_1(t)$] de operaciones marcado como DF_1 y el conjunto marcado como DF_2, se verifica que

(15) Es decir, que no está marcada como operación en incumplimiento.

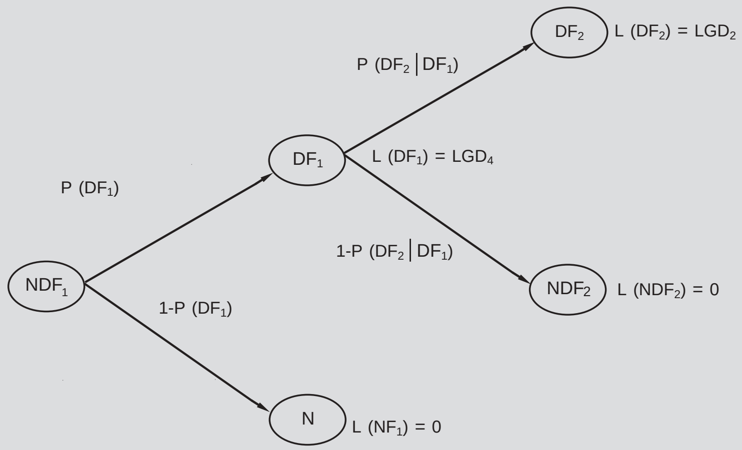
(16) Porcentaje calculado dividiendo la pérdida en unidades monetarias descontada al momento del incumplimiento entre la exposición en dicho instante.

(17) Dos operaciones se consideran similares en este contexto si están incluidas en el mismo segmento de la cartera.

(18) Operaciones que se analizan detalladamente para calcular la pérdida económica asociada, a partir de las cuales se va a estimar la E(LGD).

GRÁFICO 1

ESTADOS FINALES POSIBLES DE UNA OPERACIÓN NORMAL EN t=0



$ODF_1(t) \supseteq ODF_2(t)$. Sean n+m y m el número total de elementos de $ODF_1(t)$ y $ODF_2(t)$ de los que se conoce la pérdida, respectivamente. Para cualquier operación k (de las m posibles) ya terminada en el instante t, clasificada inicialmente como (DF_1, t_1) y finalmente como (DF_2, t_2), llamando $\Delta k=t_2-t_1$, las pérdidas en t_1 y t_2 se pueden escribir como:

$$\begin{aligned} L(k,t_1) &= LGD_1(k) * Exp(k,t_1) \\ L(k,t_2) &= LGD_2(k) * Exp(k,t_2) = LGD_2(k) * Exp(k,t_1) * e^{r(k)*\Delta k} \end{aligned} \quad [1]$$

Si todos los cobros y pagos entre t_1 y t_2 están tenidos en cuenta en [1] para el cálculo de la exposición en t_2, $L(k,t_1)$ debe ser igual a $L(k,t_2)$ descontada a t_1, con lo que se obtiene:

$$LGD_1(k) = LGD_2(k) * e^{(r(k)-i)*\Delta k} = LGD_2(k) * f(k) \quad [2]$$

Si, además, se tiene que las operaciones que dan lugar a pérdidas aparecen finalmente marcadas como DF_2, para cualquier operación (de las n existentes) ya terminada en el instante t, clasificada inicialmente como (DF_1,t_1) y que no figura como DF_2 en t, se tiene $LGD_1(k)=0$ y no hay observación de $LGD_2(k)$. Si se estiman $E(LGD_1)$ y $E(LGD_2*f)$ mediante las medias simples muestrales y se supone que $E(LGD_2*f)= E(LGD_2)*E(f)$, se encuentra que:

$$E(LGD_1) \cong E(LGD_2 * f) * \frac{n}{n+m} \cong E(LGD_2) * P(DF_2 |_{DF_1}) * E(f) \quad [2.1]$$

En el caso particular en el que $t_1=t_2$ [y, por lo tanto, $Exp(t_1)=Exp(t_2)$], $f(k)=1$ y:

$$E(LGD_1) \cong E(LGD_2) * P(DF_2 |_{DF_1}) \quad [3]$$

Si t_1 no es igual a t_2, pero se supone que para cada k, r(k)=i, se sigue verificando que f(k)=1 y, por lo tanto, también en este caso vale la fórmula [3].

Así se concluye que, para transformar las medias de las LGD correspondientes a diferentes definiciones de incumplimiento, se necesita estimar las probabilidades condicionadas asociadas (19), $P(DF_2|DF_1)$ (factor multiplicador en la ecuación [3]).

2.3. Definición de pérdida

Para calcular la pérdida L^k (asociada a la operación k-ésima), lo primero es determinar la fecha de incumplimiento (F_d) y el importe de la deuda en ese momento (D^k). Seguidamente, hay que descontar a la fecha de incumplimiento cada uno de los cobros por recuperaciones (R_i^k) y los pagos por gastos asociados a la operación (P_j^k), sumando los cobros y restando los pagos. Por último, se hace el cociente entre dicho importe neto y la deuda en el momento del incumplimiento. La pérdida, en tanto por uno, asociada a la operación k-ésima está definida por la ecuación:

$$L^k = Max\left(1 - \frac{R_i^k - P_j^k}{D^k},\ 0\right) \qquad [4]$$

Nótese que se impone que L^k sea siempre positiva o nula. La principal razón para este tratamiento es evitar la existencia de recuperaciones muy elevadas, debido a circunstancias excepcionales, que, al compensar las pérdidas ordinarias de las operaciones, podrían desvirtuar el resultado final.

2.3.1. Tipo de actualización

En la elección del tipo de actualización no hay, por el momento, un criterio claramente aceptado en la práctica. Las diferentes opciones pueden clasificarse en:

- Históricas (o estáticas): Basadas en algún tipo de interés ligado a la operación en una fecha concreta. Dicho tipo se utiliza para calcular la pérdida, con independencia de la fecha en la que se efectúe dicho cálculo. Son ejemplos de esta categoría: la elección del tipo de concesión de la operación, el tipo de demora aplicado a la operación en el momento del incumplimiento, algún tipo externo (por ejemplo, interbancario a un año) existente en el momento de la operación.

- Dinámicas: Se trata de asociar un tipo de interés conocido en el momento de efectuar el cálculo de la LGD, con la idea de incorporar

(19) En el caso general, en el que operaciones que en el instante t están marcadas como DF_1 y no marcadas como DF_2 puedan terminar en pérdidas, la relación es ligeramente más complicada.

la mejor información disponible a la hora de estimar las pérdidas. Como no hace falta que el tipo aplicado a una operación sea conocido en el momento del incumplimiento, este tipo de opción permite utilizar la estructura de tipos de interés conocida en el momento del cálculo de la LGD para descontar cada flujo asociado a la recuperación de la operación a la tasa adecuada al período transcurrido entre el instante del incumplimiento y el flujo correspondiente. Como ejemplos de esta clase destacan: el uso de una estructura de tipos [por ejemplo, de *swaps* de tipos de interés (IRS)] en el momento en el que se está efectuando el cálculo y el uso de un tipo fijo determinado en el momento del cálculo (por ejemplo, el interbancario a un año).

Una discusión detallada de los diferentes métodos, sus ventajas e inconvenientes, y de su impacto en el cálculo de la pérdida para dos carteras de operaciones hipotecarias en España, puede encontrarse en Moral y Oroz (2002).

En principio, cuanto menores sean los plazos de recuperación, la elección del tipo de actualización debería tener menor impacto en el cálculo de la pérdida. Como se comenta al hablar de las adjudicaciones (20), se ha considerado que la recuperación de la operación crediticia finaliza en el momento de la subasta con venta a terceros o adjudicación del inmueble. Este criterio acorta el plazo de recuperación y atenúa el impacto directo del tipo de interés utilizado. En este trabajo se ha optado por utilizar una opción dinámica basada en un único tipo de actualización, r, para todos los flujos. Este parámetro r debe fijarse teniendo en cuenta la finalidad para la que se calcula la LGD. Por ejemplo, si se quiere estimar la LGD con la finalidad de calcular pérdidas esperadas en el horizonte de un año, no parece adecuado utilizar tipos ligados al momento en el que produjo el incumplimiento. Si el plazo medio de recuperación estimado es de 1,5 años, el tipo adecuado podría ser un tipo medio entre los tipos a dos y tres años, en el momento de hacer la estimación.

2.4. Gastos imputados y modelo de imputación de gastos

Desde el momento en que una entidad decide emprender un procedimiento judicial hipotecario, hasta que finalmente se adjudica (a la propia entidad o a un tercero) el inmueble en garantía, se produce una serie de gastos [judiciales e ITP (21), básicamente] de los que, normalmente, no es posible obtener su importe con la información mecanizada disponible. También es muy difícil obtener su importe exacto a partir de los expedientes existentes en las entidades, pero el estudio y análisis de casos representativos en los que tal detalle sí está disponible permiten estimar los citados costes.

Con la información obtenida de las operaciones analizadas, se ha construido un modelo de imputación de gastos para este tipo de prés-

(20) Apartado 2.5, Tratamiento de las adjudicaciones de bienes.

(21) ITP, impuesto sobre transmisiones patrimoniales que tiene que pagar quien se adjudica el bien.

tamos hipotecarios, estimando los gastos de representación procesal y los gastos de adjudicación. Las principales ventajas de disponer de un modelo de imputación de gastos son:

- Permite incorporar gastos a las operaciones de las que no se dispone del detalle de los gastos reales.
- Aun cuando se disponga de los gastos reales históricos, esa información puede ser bastante heterogénea, ya que se trata de datos obtenidos en diferentes momentos del tiempo, con estructuras de costes (propios y ajenos) posiblemente diferentes. Con un modelo, se puede homogeneizar, parcialmente, esta información.
- Hace posible utilizar la información disponible (incorporando cambios en las estructuras de costes judiciales, cambios en los impuestos, etc), para obtener los gastos de las operaciones pasadas en las condiciones actuales. De esta forma, se aprovechan todos los datos históricos (que suelen ser escasos y pueden abarcar períodos largos).

2.4.1. Gastos judiciales

Se corresponden con los honorarios —IVA incluido— tanto del letrado como del procurador, así como una serie de gastos necesarios para el inicio y para el desarrollo del proceso (requerimientos de pago al deudor, certificaciones del Registro de Propiedad, edictos que se publican en el BOE y/o en el BOP, etc.).

Los honorarios de letrado y procurador suelen girar en torno al importe de la deuda reclamada (22) (DR), que generalmente no coincide con la cantidad en incumplimiento (D). Para no complicar excesivamente el modelo, se han expresado dichos gastos en función de D. A efectos de su inclusión en la fórmula de cálculo de la pérdida, se supone que se pagan en la mitad del período entre la fecha de incumplimiento (F_d) y la fecha de recuperación (F_r). Su expresión viene dada por:

$$P_1^k = g_1 D^k (1+r)^{-\frac{(Fr^k - Fd^k)}{365*2}} \quad [5]$$

En cada caso concreto, habrá que proceder a la correspondiente estimación de g_1 a partir de la información disponible.

2.4.2. Gastos de adjudicación del bien

Si el proceso judicial termina —como es frecuente— en la adjudicación del bien a la entidad ejecutante (o de una instrumental de su mismo

(22) Que incluye intereses de demora y otros gastos que se repercuten al deudor que no están incluidos en el importe en incumplimiento.

grupo económico), esta incurre en una serie de gastos adicionales, que en España, básicamente, son los originados por la transmisión de la propiedad (pago del ITP y escritura pública, e inscripción de la misma en el Registro de la Propiedad) y por la cancelación de la hipoteca ejecutada.

Los más significativos son los derivados de la transmisión del inmueble, y pueden establecerse en torno al 7 % del valor de adjudicación del bien [6 % de ITP (23) y 1 % de escrituración pública e inscripción registral]. Nuevamente, se expresan dichos gastos en función de la deuda en el momento de incumplimiento D. Dicha relación debe obtenerse a partir del estudio y verificación de la política de adjudicaciones de la entidad y del análisis de los costes de adjudicación históricos. A efectos de su inclusión en la fórmula de cálculo de la pérdida, se ha supuesto que se pagan en la fecha de adjudicación (F_a). Su expresión viene dada por:

$$P_2^k = g_2 D^k \ (1+r)^{-\frac{(Fa^k - Fd^k)}{365}} \qquad [6]$$

2.5. Tratamiento de las adjudicaciones de bienes

Otro problema es decidir cuándo se considera recuperada una cantidad y en qué momento termina el proceso de recuperación. En el caso de las entidades de mayor tamaño, es muy frecuente que la adjudicación se haga a una sociedad vinculada (por ejemplo, una filial 100 % del grupo, que después gestiona la cartera de inmuebles adjudicados). Es fundamental analizar detalladamente la política de adjudicaciones de la entidad. Frecuentemente, estas políticas (24) fijan reglas para los precios máximos en las adjudicaciones y precios mínimos para las ventas o adjudicaciones a terceros. En la cartera analizada, la regla general era que el grupo se adjudicaba el bien, como máximo, por un α % (<100 %) del mínimo entre el valor de tasación y el valor de la deuda reclamada.

Desde un punto de vista teórico, lo ideal sería poder incluir los datos de venta a terceros y, de esta manera, calcular pérdidas finales para el grupo. Este enfoque no es adecuado, debido a la necesidad de incluir no solo los resultados de las ventas que ya se han producido a terceros, sino también las pérdidas latentes en la cartera de inmuebles adjudicados. Además, también se necesitaría conocer los gastos de mantenimiento, vigilancia, etc., directamente imputables a dicha cartera.

Otra solución más realista, y suficiente en muchos casos, es tratar de limitar el alcance, dando por concluido el proceso de recuperación, a efectos de estos cálculos, en el momento de la adjudicación del bien. La metodología desarrollada trata el importe por el que se adjudica al grupo el bien, minorado por un determinado coeficiente reductor (inten-

(23) Al tratarse de un impuesto autonómico, en algunas Comunidades el tipo es actualmente del 7 % (por ejemplo, Madrid). El anejo 4 recoge los diferentes tipos existentes a la fecha de redacción de este documento en las distintas Comunidades Autónomas.

(24) Que, para mayor complejidad, pueden cambiar con el tiempo.

tando reflejar que, incluso en el mejor de los casos, habrá unos gastos de mantenimiento, de reparación y financieros hasta el momento de la venta a terceros), como una cantidad recuperada. El coeficiente reductor (d) aplicado debe tener en cuenta la política de adjudicaciones de la entidad, la calidad de las tasaciones, los plazos de venta fuera del grupo, los resultados de las ventas ya realizadas fuera del grupo, etc. Por lo tanto, se observa que dicho coeficiente cambia con el tiempo. El valor utilizado debe estimarse en el momento de realización del estudio. Este enfoque permite separar los resultados directamente imputables a la inversión crediticia, de los derivados de la existencia de una cartera de inmuebles adjudicados.

El cálculo del flujo asociado a la operación k-ésima mediante la adjudicación al grupo de un bien por un importe A_i en la fecha F_a, viene dado por:

$$R_i^k = d\,A_i^k\,(1+r)^{-\frac{(Fa^k - Fd^k)}{365}} \qquad [7]$$

2.6. Expresión general de la pérdida asociada a una operación

Combinando las expresiones de las fórmulas [4], [5], [6] y [7] se obtiene la expresión general para la pérdida asociada a la operación k-ésima en función de los parámetros del modelo:

$$L^k(d, g_1, g_2, r) = Max\left(1 - \frac{\sum R_i^k - \sum P_j^k}{D^k},\ 0\right) =$$

$$Max\left(1 - \frac{C^k(r)}{D^k} - d\,\frac{A^k}{D^k}(1+r)^{-\frac{(Fa^k - Fd^k)}{365}} + g_1\,(1+r)^{-\frac{(Fr^k - Fd^k)}{365*2}} + sign(A^k)\,g_2\,(1+r)^{-\frac{(Fa^k - Fd^k)}{365}},\ 0\right) \qquad [4.1]$$

donde A^k es el valor por el que el bien se ha adjudicado al grupo (25) y $C^k(r)$ es el neto de los valores descontados del resto de cobros y pagos habidos durante el proceso de recuperación de dicha operación. Se tiene que, fijados los coeficientes asociados a la operación y el valor de r, la función anterior es localmente lineal en los parámetros (d, g_1, g_2), excepto en los puntos del hiperplano H^k:

$$H^k \equiv 1 - \frac{C^k(r)}{D^k} - d\,\frac{A^k}{D^k}(1+r)^{-\frac{(Fa^k - Fd^k)}{365}} + g_1\,(1+r)^{-\frac{(Fr^k - Fd^k)}{365*2}} + sign(A^k)\,g_2\,(1+r)^{-\frac{(Fa^k - Fd^k)}{365}} = 0 \qquad [8]$$

Por comodidad, en adelante se utilizará para la expresión del hiperplano H^k, asociado a la operación k-ésima, la nomenclatura reducida:

$$H^k \equiv p^k(d, g_1, g_2) = 0 \qquad [8.1]$$

(25) Si no ha habido tal adjudicación, dicho valor es cero.

En el subespacio R^{k+}, del espacio de parámetros, definido por $p^k(d,g_1,g_2)>0$, se tiene:

$$L^k\left(d,g_1,g_2\right)=p^k\left(d,g_1,g_2\right) \qquad [4.2]$$

Finalmente, en el subespacio R^{k-}, del espacio de parámetros, definido por $p^k(d,g_1,g_2)<0$, la pérdida es cero.

3. POBLACIÓN, MUESTRA ANALIZADA Y ESTIMADORES

3.1. Población

Se parte de una población de N operaciones, formada por préstamos hipotecarios concedidos a personas físicas hasta la fecha t, desde 5 años antes. Se denomina O(t) a ese conjunto de operaciones.

Se trata de un conjunto de operaciones bastante heterogéneo en cuanto a los bienes hipotecados, sin que la codificación de las garantías existente en la base de datos de trabajo permita segmentar las operaciones por tipos de bienes (por ejemplo, primera vivienda, segunda residencia, local comercial).

Por importes concedidos, O(t) tiene una media de casi 8 millones de pesetas. El 60 % de las operaciones se concedió con importe entre 5 y 15 millones, siendo inferiores a 15 millones más del 90 % de ellas. La mayor parte de las operaciones se concedió con un LTV inferior al 80 %.

Se llama $ODF_1(t)$ al conjunto de operaciones de O(t) marcadas como DF_1 en la fecha de análisis t. El cardinal de $ODF_1(t)$ se representa por nd_1. Análogamente, se definen ODF_2 y nd_2.

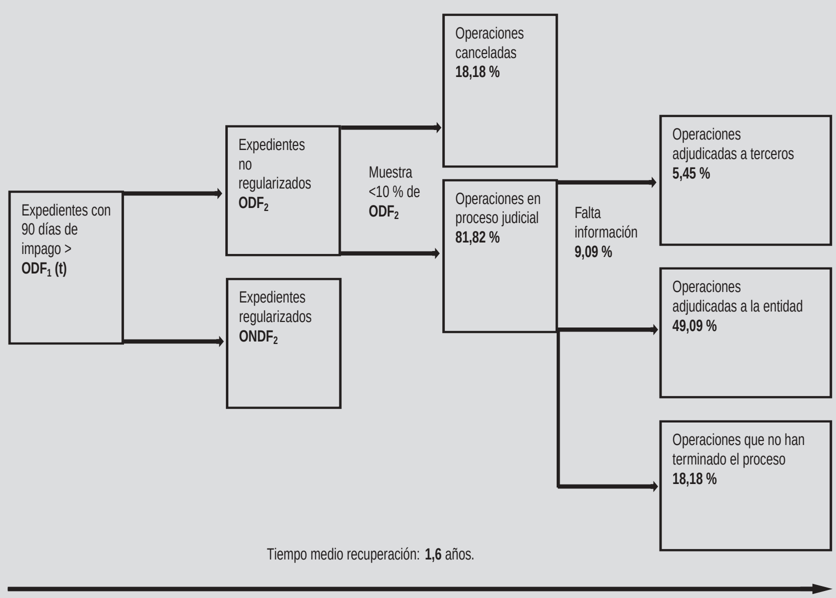
3.2. Muestra analizada

Dados los costes asociados a analizar los expedientes (tanto por la dificultad que supone recuperar expedientes relativamente antiguos y geográficamente dispersos como por el tiempo que consume revisarlos para determinar la pérdida económica de las operaciones), se ha trabajado con una muestra pequeña, que representa menos del 5 % de los expedientes clasificados como DF_1. El gráfico 2 esquematiza la relación entre población y muestra.

Tras el análisis de la política de recuperaciones, se ha utilizado la hipótesis de que, en las operaciones recuperadas clasificadas como NDF_2 (operaciones regularizadas mediante pago en efectivo), la pérdida es cero. En consecuencia, se ha centrado la revisión en los nd_2 expedientes

GRÁFICO 2

RELACIÓN ENTRE LA POBLACIÓN Y LA MUESTRA



no regularizados, de los que se ha extraído una muestra aleatoria de 55 expedientes (menos del 10 % de nd_2). Al revisar estos expedientes, se encontraron 10 que se habían regularizado en efectivo con pérdida económica cero. De los 45 restantes, 10 no habían concluido el proceso judicial (se han tratado suponiendo que su LGD es la media del resto de la muestra en el cálculo de la pérdida), y de 5 no se ha podido encontrar información que permita calcular la pérdida (se tratan igual que los no terminados). Por último, de los 30 expedientes que habían concluido el proceso de recuperación judicial, solo 3 se habían adjudicado inicialmente a terceros.

3.3. Cálculo de la LGD muestral y tratamiento de los elementos sin información completa

A la hora de calcular la LGD muestral, hay que decidir el tratamiento de los elementos sin información completa. Se trata de operaciones seleccionadas para formar parte de la muestra, en las que, tras la recopilación de toda la información disponible (básicamente, contenida en el correspondiente expediente), finalmente no es posible determinar (sin incertidumbre significativa) su pérdida. Las principales causas que explican que esta situación se presente son: la existencia de operaciones relativamente antiguas en las que se ha extraviado información del expediente y la de operaciones que están aún en el proceso de recuperación en el momento del estudio.

En principio, los tres candidatos para estimar la $E(LGD_2)$, a partir de los elementos con información completa (LGD_m), son:

- La media aritmética simple muestral de los L^k: es la elección natural. Si los importes en incumplimiento están depurados de casos extremos (sobre todo, exposiciones muy pequeñas), esta elección produce las estimaciones de $E(LGD_2)$ más estables ante variaciones en la muestra.

- La media aritmética de los L^k, ponderada con los importes en incumplimiento D^k: parece atractivo explotar la posible correlación existente entre importes en incumplimiento y L^k en la muestra (26). El problema es que, con tamaños muestrales pequeños, aumenta la inestabilidad del proceso de estimación de la LGD. Además, como se verá en el apartado correspondiente, con la muestra utilizada empeora la estimación por intervalos.

- La mediana muestral: la idea es utilizar un estadístico que no sea sensible a valores extremos, como forma de atenuar la fuerte variabilidad muestral con tamaños de muestra pequeños. Un problema es que, dependiendo de cómo sea la distribución empírica, puede conducir a estimaciones absurdas (27).

En el caso que se presenta como ejemplo, a partir de la muestra de tamaño 55, la pérdida se pudo calcular en solo 40 casos (i=1,…,40). Denominando L^k a la pérdida correspondiente a cada una de las operaciones con información completa contenidas en la muestra, se obtiene:

$$LGD_{m,40} = \frac{1}{40}\sum_{k=1}^{40} L^k \qquad [9.1]$$

$$LGD_{mp,40} = \frac{\sum_{k=1}^{40} L^k * E^k}{\sum_{k=1}^{40} E^k} \qquad [9.2]$$

$$LGD_{med,40} = Mediana\left\{L^1, ..., L^{40}\right\} \qquad [9.3]$$

Hay que decidir el tratamiento de los elementos con información incompleta. En principio, las posibilidades son:

- Ignorarlos: es lo más sencillo, pero no parece aconsejable, debido a que, precisamente, parte de estos elementos (28) son los que pro-

(26) En el apartado 2.2 se ha supuesto que no existe correlación entre las exposiciones y la LGD.
(27) Por ejemplo, en la distribución de pérdidas asociada a los DF_1 la mediana es cero.
(28) En concreto, los elementos que no han terminado el proceso de recuperación.

bablemente se parecen más a los que generarán las próximas pérdidas. Si en la muestra hay un porcentaje significativo de elementos sin información completa, ignorarlos puede introducir sesgos importantes.

- Tratarlos como una función del resto de valores: por ejemplo, la media muestral de los elementos con información completa multiplicada por un determinado coeficiente, que tiene en cuenta el resto de información disponible específica de cada expediente.

- Considerarlos variables aleatorias, en las que los parámetros dependen de estadísticos muestrales de los elementos con información completa y del resto de información disponible. Es el método más general; su utilización puede limitarse al caso en el que se desee estimar la distribución del estimador o la sensibilidad a variaciones en la muestra.

En los apartados que siguen, a las operaciones con información incompleta (k=41,…,55, en las que no ha sido posible calcular la L^k) se les ha asociado como pérdida porcentual la obtenida a partir de las correspondientes fórmulas anteriores, mediante la expresión: $L^k=\beta * \underline{L}_{,40}$.

La principal razón para modelizar de esta forma los expedientes sin información es que permite utilizar la información previa que pudiera existir sobre la calidad de los expedientes cuyo proceso de recuperación no ha terminado, o aquellos en los que no se dispone de información suficiente para determinar la pérdida. Por ejemplo, si se piensa que los expedientes de las operaciones no terminadas en el momento de análisis son peores que los ya terminados, se puede utilizar un valor de $\beta>1$.

En este caso particular, se ha utilizado $\beta=1$, ya que no se tiene evidencia de que dichos expedientes sean de calidad diferente al resto de los analizados. Así, finalmente, incluyendo todas las operaciones, las fórmulas anteriores se transforman en:

$$LGD_m = \frac{1}{55}\sum_{k=1}^{55} L^k = \frac{1}{40}\sum_{k=1}^{40} L^k \qquad [10.1]$$

$$LGD_{mp} = \frac{\sum_{k=1}^{55} L^k * E^k}{\sum_{k=1}^{55} E^k} = \frac{\sum_{k=1}^{40} L^k * E^k}{\sum_{k=1}^{40} E^k} \qquad [10.2]$$

$$LGD_{med} = Mediana\left\{L^1,...,L^{40},...,L^{55}\right\} = Mediana\left\{L^1,...,L^{40}\right\} \qquad [10.3]$$

Aun cuando los valores numéricos (en este caso, con $\beta=1$) no cambian, es importante notar que se están considerando muestras de tamaño 55 y no de tamaño 40. Lo anterior es especialmente importante en el estudio de la estabilidad ante variaciones en la muestra.

Combinando las ecuaciones [10.1] y [4.1], se obtienen los estimadores de la LGD muestral para cada uno de los casos. Por ejemplo, para la media muestral se llega a:

$$LGD_m(d,g_1,g_2)=\frac{1}{55}\sum_{k=1}^{55}L^k(d,g_1,g_2)=\quad\frac{1}{40}\sum_{k=1}^{40}L^k(d,g_1,g_2)\qquad [11.1]$$

Para cada elección de los parámetros (d^0,g_1^0,g_2^0), sea $G(d^0,g_1^0,g_2^0)$ el subconjunto de {1,…,40, 41,…,55} que recoge los índices k de los hiperplanos $H^k \equiv p^k(d,g_1,g_2)=0$, tales que $p^k(d^0,g_1^0,g_2^0)>0$. Supuesto que $p^k(d^0,g_1^0,g_2^0)\neq 0$ para todo k, ese conjunto de índices permanece fijo ante variaciones en los valores de los parámetros, mientras no se llegue a algún punto de los hiperplanos H^k. Se denomina $R[G(d^0,g_1^0,g_2^0)]$ al conjunto de puntos del espacio de parámetros que tienen asociados el mismo subconjunto de índices $G(d^0,g_1^0,g_2^0)$. Entonces, para cualquier punto $(d, g_1, g_2) \in R[G(d^0,g_1^0,g_2^0)]$, la ecuación [11.1] se puede escribir como:

$$LGD_m(d,g_1,g_2)=\frac{1}{55}\sum_{k\in G}LGD^k(d,g_1,g_2)=\frac{1}{55}\sum_{k\in G}p^k(d,g_1,g_2)\qquad [12.1]$$

Análogamente, se obtienen las correspondientes expresiones para la media ponderada y para la mediana, válidas para puntos $(d, g_1, g_2) \in R(G(d^0,g_1^0,g_2^0))$:

$$LGD_{mp}\left(d,g_1,g_2\right)=\frac{\sum_{k\in G}p^k(d,g_1,g_2)*E^k}{\sum_{k=1}^{55}E^k}\qquad [12.2]$$

$$LGD_{med}\left(d,g_1,g_2\right)=Mediana\left\{p^k\left(d,g_1,g_2\right),\quad k\in G(d^0,g_1^0,g_2^0)\right\}\qquad [12.3]$$

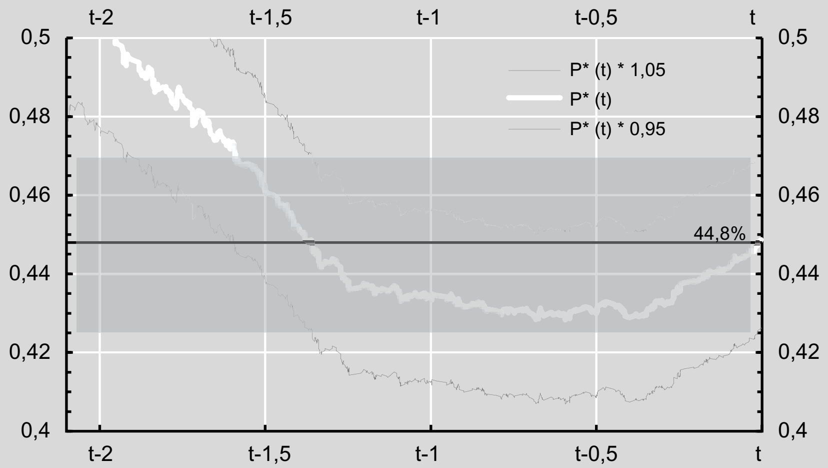
3.4. Estimación de la LGD en la población

El procedimiento propuesto de estimación de la LGD en la población, a partir de los valores muestrales, es el siguiente. En primer lugar, se utiliza LGD_m como estimador de la $E(LGD_2)$ asociada a las operaciones en incumplimiento no regularizadas (LGD asociada a la definición de incumplimiento DF_2). A continuación, aplicando la fórmula [3], se calcula la estimación correspondiente a $E(LGD_1)$. Para aplicar la fórmula anterior, se necesita un estimador del multiplicador ($P(DF_2|_{DF1})$) en la fecha de análisis. El estimador más sencillo es el cociente entre el número de incumplimientos DF_2, desde el origen de la base de datos hasta el momento t, ($dn_2(0,t)$), y el número de incumplimientos DF_1, ($nd_1(0,t)$), esto es:

$$P(DF_2/DF_1)\cong\frac{nd_2(0,t)}{nd_1(0,t)}=P^*(t)\qquad [13]$$

GRÁFICO 3

EVOLUCIÓN DEL ESTIMADOR DE LA PROBABILIDAD CONDICIONADA



Con lo que, en este caso concreto, el estimador puede calcularse con la fórmula siguiente:

$$E(LGD^{p}) \cong P(DF_2/DF_1)*LGD_m \cong \frac{nd_2}{nd_1}LGD_m = 0{,}4484*LGD_m \quad [14]$$

Esta regla de conversión entre la E(LGD^p) con una definición de incumplimiento DF_1 (90 días de impago) y la E(LGD) asociada a la otra definición de incumplimiento más restrictiva DF_2 (excluyendo las operaciones que finalmente han regularizado su situación mediante pago en efectivo) puede ir cambiando con el tiempo y, por lo tanto, tiene que recalcularse periódicamente. Es fundamental estudiar la estabilidad a lo largo del tiempo del estimador de la probabilidad condicionada. En el gráfico 3 se presenta la evolución (durante dos años) del factor de conversión en el ejemplo considerado.

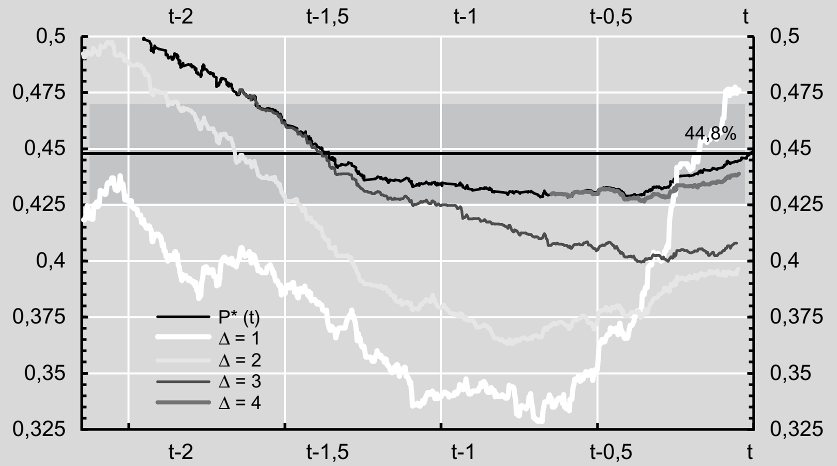
En el eje de abscisas, t representa el instante en el que se hizo la estimación del factor de conversión. Se observa que dicho factor está comprendido en un intervalo de ±5 % (región sombreada del gráfico 3) de la estimación utilizada (44,8 %), durante más de año y medio (29). Las líneas más finas representan, para cada instante t, los extremos de los intervalos [0,95P^*(t), 1,05P^*(t)]. Gráficamente, se confirma que es equivalente encontrar la región en la que el intervalo centrado en el 44,8 % contiene los valores de P^*(t), a determinar los valores de t para los que el intervalo centrado en P^*(t) contiene el 44,8 %.

Se pueden utilizar otros estimadores de la probabilidad condicionada. Los más naturales se corresponden con fórmulas similares a [13], en las

(29) Dicho cálculo está realizado con la información disponible en la base de datos utilizada que solo tenía almacenada la fecha correspondiente a la entrada en incumplimiento con la definición DF_1.

GRÁFICO 4

COMPARACIÓN ENTRE DIFERENTES ESTIMADORES DE $P(DF_2 \mid DF_1)$



que las operaciones consideradas no son todas las existentes en la base histórica, sino solo las incluidas en los intervalos ($t–\Delta$, t). El gráfico 4 muestra la comparación entre el estimador utilizado de la probabilidad condicionada y los estimadores basados en utilizar solamente el conjunto de operaciones existentes durante un período de duración Δ, para valores de Δ=un año, dos años, …, tres años y cuatro años.

Se ve que, si se considera un período muy largo (Δ= cuatro años), el estimador es muy parecido al anteriormente utilizado. Con plazos cortos (Δ= un año), los estimadores son muy inestables. La decisión de cuál es el estimador óptimo depende de la información disponible y de la finalidad perseguida. Aquí, de cara a obtener una estimación de E(LGD) en la población, se ha preferido utilizar el estimador basado en toda la información histórica, al igual que se ha hecho al estimar $E(LGD_2)$.

3.5. Estimación puntual de la $E(LGD_1)$

A partir de la muestra analizada y del resto de información disponible, se han estimado los parámetros (d, g_1, g_2) necesarios para el cálculo de la pérdida en el ejemplo considerado.

3.5.1. Coeficiente de gastos judiciales, g_1

Partiendo de una estimación del 9 % de la deuda reclamada (8 % para el letrado y 1 % para el procurador) más un 1 % para el resto de gastos procesales, se ha obtenido finalmente un coeficiente (30) g_1=10 %. Esta valoración de los honorarios del letrado guarda, además, bastante rela-

(30) Como DR>D, el 10 % es una estimación a la baja de estos gastos.

ción con las tarifas orientativas que, para los importes en cuestión, tienen publicadas los distintos Colegios de Abogados de España.

3.5.2. Coeficiente de gastos de adjudicación, g_2

Después de analizar la política de adjudicaciones, y con la información histórica disponible, se han estimado los gastos de adjudicación en términos de la deuda en el momento del incumplimiento, como 6,5 % [que es, aproximadamente, el α % (31) del ITP medio (32) de su cartera]. Los gastos originados por la cancelación de hipoteca se estiman como el 1,5 % de la deuda. En conjunto, implican un coeficiente de gastos de adjudicación g_2 del 8 %.

3.5.3. Coeficiente reductor por adjudicaciones propias, d

En el caso concreto analizado, dada la imposibilidad de obtener datos que permitieran la estimación estadística del parámetro d, se ha optado por utilizar d=90 %, factor que, dada la política de precios máximos para las adjudicaciones, y basándose en el análisis concreto de ciertos casos y en experiencias previas, se considera razonable.

De la muestra analizada, con los valores para los parámetros del modelo d=90 %, g_1=10 % y g_2=8 %, aplicando las fórmulas [10.1], [10.2], [10.3], se obtiene: LGD_m=28,20 %, LGD_{mp}=34,21 %, LGD_{med}=25,75 %. Aplicando la ecuación [14], las estimaciones de $E(LGD^p)$ para la población son: LGD^p_m =12,65 %, LGD^p_{mp}=15,34 % y LGD^p_{med}=11,55 %. Los valores anteriores incluyen la imputación de gastos y el tratamiento especial de las recuperaciones vía adjudicación al propio grupo que ya se han comentado anteriormente. Su impacto es muy elevado, como pone de manifiesto que, por ejemplo, las correspondientes cifras, sin incluir imputación de gastos ni aplicación de coeficiente reductor a los flujos por adjudicaciones, para la estimación basada en la media simple, son el 14,98 % para la muestra, y el 6,72 % para la estimación poblacional.

La dispersión de los valores muestrales de la pérdida es muy alta, con una desviación típica (33) del 26,12 %. Dado el reducido tamaño muestral, la elevada varianza existente en la muestra, y el tratamiento de los expedientes sin información completa, los métodos de inferencia estadística basados en fórmulas cerradas no son, en principio, adecuados para analizar la confianza de la estimación de la LGD. Sí se puede analizar la estabilidad del proceso de estimación anterior ante variaciones en la muestra (*Jacknife*) y dar intervalos de confianza para la LGD de población, basados en procesos de *Bootstrap*.

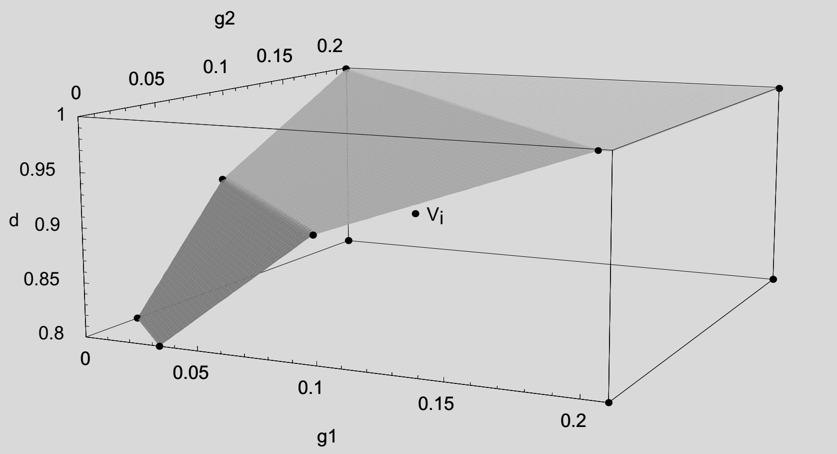
(31) Se está utilizando el conocimiento de la política de adjudicaciones, que en este caso era, en principio, no pujar por encima del α % del valor de la deuda, además de considerar el valor de tasación y las cargas del bien.

(32) Ya que los tipos pueden variar entre el 7 %, 6 % y 4 %, según Comunidades Autónomas y tipos de bienes.

(33) Sin considerar los elementos sin información.

GRÁFICO 5

REGIÓN ASOCIADA AL PUNTO V_i = (d = 90 %, g_1 = 10 %, g_2 = 8 %)



3.5.4. Sensibilidad a los parámetros de imputación de gastos

Es claro que una revisión con mejor información en la base de datos y más expedientes analizados permitiría obtener valores más precisos para los coeficientes de gastos. En este apartado se va a analizar la sensibilidad de LGD_m y LGD^p frente a los parámetros de imputación de gastos jurídicos (g_1), gastos de adjudicación (g_2) y coeficiente reductor aplicado a los flujos por recuperaciones intergrupo (d). Tomando como punto de partida del espacio de parámetros el utilizado en la estimación puntual $V_i=(d^0, g_1^0, g_2^0)=(0{,}9,\ 0{,}1,\ 0{,}08)$, se obtiene el conjunto $R[G(d^0, g_1^0, g_2^0)]$= puntos pertenecientes al cubo unidad situados por debajo de las caras sombreadas del gráfico 5:

Las caras sombreadas pertenecen a hiperplanos, H^k, en los que el semiespacio negativo contiene el punto V_i; es decir, en los que $p^k(V_i)<0$

Con el estimador de $E(LGD_2)$ basado en la media simple y utilizando la ecuación [12.1], se obtiene:

$$LGD_m = 69{,}31\,\% + 0{,}5518 g_1 + 0{,}4848 g_2 - 0{,}5612 d \qquad [15]$$

En efecto, para los valores g_1=10 %, g_2=8 % y d=90 % se tiene un valor para LGD_m del 28,20 %. La estimación de la LGD_p (d, g_1, g_2), utilizando las ecuaciones [14] y [15], es:

$$LGD_m^p = 31{,}08\,\% + 0{,}2475 g_1 + 0{,}2174 g_2 - 0{,}2516 d \qquad [16]$$

Para los puntos situados por encima de las caras sombreadas, el conjunto de índices R varía y, por lo tanto, al aplicar las fórmulas [12.1] y [14] se obtendrían otras expresiones válidas para sus correspondientes conjuntos R(). Por ejemplo, para d=100 % y $g_1=g_2=0$ %, la ecuación [16] da un importe del 6,01 %, frente al valor real del 6,72 %.

Con el estimador de $E(LGD_2)$ basado en la media ponderada, utilizando la ecuación [12.2] resulta:

$$LGD_{mp} = 76{,}69\,\% + 0{,}6526\,g_1 + 0{,}5854\,g_2 - 0{,}5965d \qquad [17]$$

Utilizando la ecuación [14], se obtiene para la población en este caso (34):

$$LGD_{mp}^{p} = 34{,}39\,\% + 0{,}2926\,g_1 + 0{,}2625\,g_2 - 0{,}2675d \qquad [18]$$

Por último, para el estimador de $E(LGD_2)$ basado en la mediana, se llega a las siguientes expresiones, válidas en un entorno del punto inicial (35) (d=90 %, g_1=10 %, g_2=8 %):

$$LGD_{med} = 105{,}45\,\% + 0{,}4866\,g_1 + 0{,}4735\,g_2 - 0{,}9817d \qquad [19]$$

Utilizando la ecuación [14], se obtiene para la población:

$$LGD_{med}^{p} = 47{,}29\,\% + 0{,}2182\,g_1 + 0{,}2123\,g_2 - 0{,}4402d \qquad [20]$$

3.5.5. Estabilidad del proceso de estimación de $E(LGD_1)$

Para contrastar la estabilidad de los resultados obtenidos frente a cambios pequeños en la muestra base, se puede estudiar cómo varían las estimaciones anteriores si se utilizan muestras parecidas a la original. La forma más sencilla de generar dichas muestras es eliminar o añadir algún elemento a la muestra original.

3.5.5.1. Jacknife *del estimador de $E(LGD_1)$*

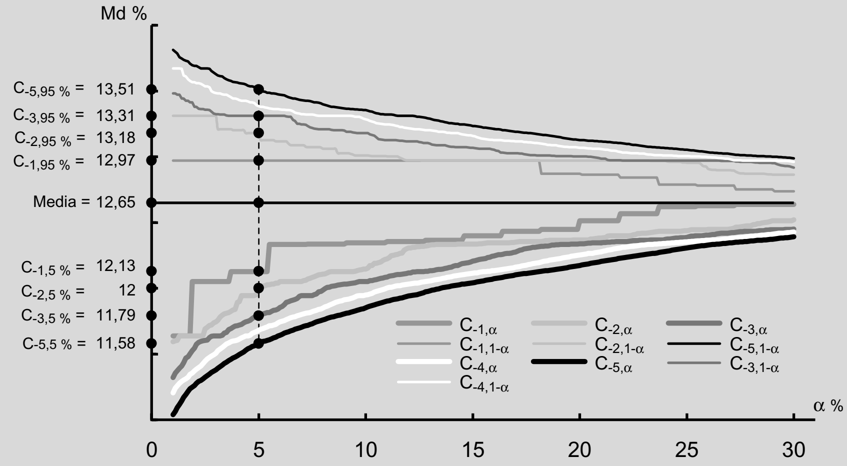
Se ha efectuado un proceso de eliminación de operaciones de la muestra *(Jacknife),* para obtener las distribuciones de la estimación de $E(LGD_1)$. En particular, se han obtenido las distribuciones asociadas a eliminar, de todas las formas posibles, un valor muestral (n=1), dos valores (n=2),…, cinco valores (n=5). El gráfico 6 resume los percentiles 5 % y 95 %, para cada uno de los casos del estimador basado en la media simple.

(34) Se supone que los pesos asociados a las operaciones regularizadas no están disponibles.

(35) En este caso, la región válida para las ecuaciones [19] y [20] es menor que la correspondiente región para las ecuaciones [15], [16], [17] y [18].

GRÁFICO 6

PERCENTILES OBTENIDOS ELIMINANDO OPERACIONES DE LA MUESTRA



Se observa que, eliminando tres puntos muestrales de todas las formas posibles, la correspondiente distribución del valor estimado de $E(LGD_1)$ alcanza el percentil correspondiente al 5 %, en $C_{3,5\ \%}$=11,79 %, y el correspondiente al 95 %, en $C_{3,95\ \%}$=13,31 %.

3.5.5.2. Bootstrap *condicional del estimador de $E(LGD_1)$*

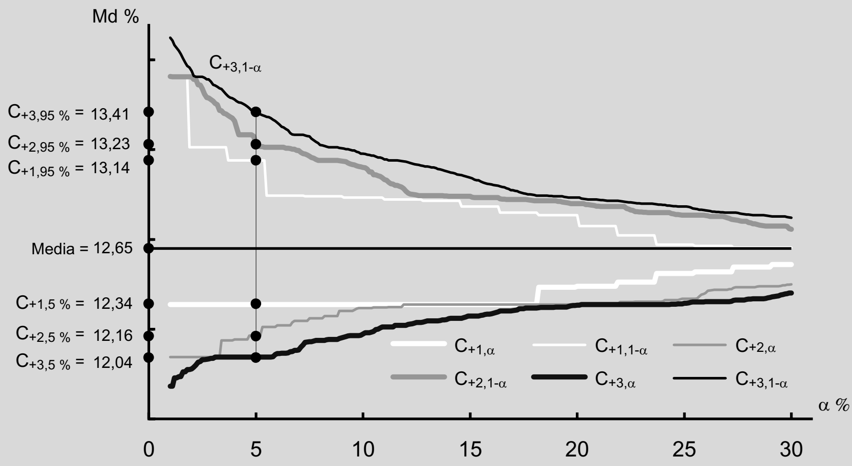
Si, en vez de eliminar, se añaden nuevos elementos, obtenidos por muestreo con reposición a partir de la muestra original, se puede analizar la estabilidad de la estimación de E(LGD) frente a incrementos muestrales. En concreto, el gráfico 7 resume los percentiles correspondientes al 5 % y al 95 %, para los casos en los que se han añadido: un elemento (n=1), dos elementos (n=2) y tres elementos (n=3) de todas las formas posibles, con el estimador de $E(LGD_1)$ basado en la media simple.

Se observa que, si se añaden a la muestra tres elementos de todas las formas posibles, se obtiene que el percentil de la estimación de LGD correspondiente al 5 % es $C_{+3,5\ \%}$=12,04 % y el correspondiente al 95 % en $C_{+3,95\ \%}$=13,41 %.

Por lo tanto, combinando ambos resultados, se puede concluir que la estabilidad del proceso de estimación frente a variaciones en la muestra es elevada, como pone de manifiesto que, para n=3 (desaparición o incremento en la muestra de tres elementos), el intervalo [11,79 %, 13,41 %] contiene el valor de la estimación de la $E(LGD_1)$ en más del 90 % de los casos. Este último intervalo, en términos de la definición de incumplimiento DF_2, se transforma en [26,3 %, 29,9 %].

GRÁFICO 7

PERCENTILES OBTENIDOS AGREGANDO ELEMENTOS A LA MUESTRA



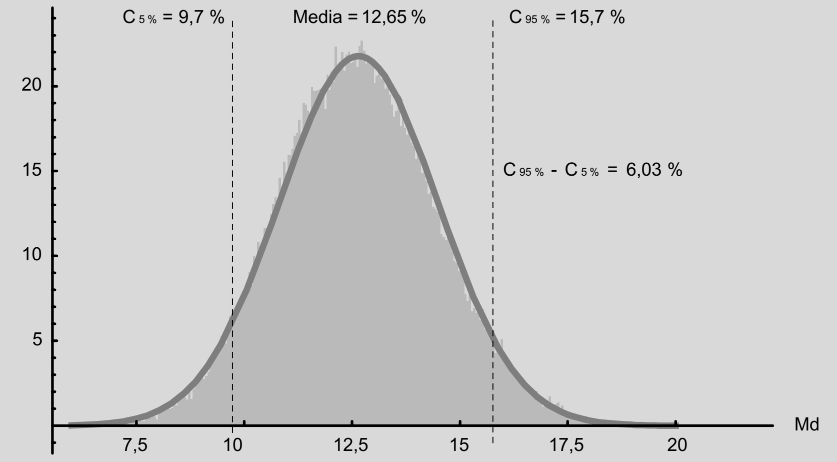
3.6. Estimación por intervalos de $E(LGD_1)$

Para estudiar la estabilidad frente a cambios grandes, los métodos combinatorios como los anteriormente utilizados no son viables, dado el número de combinaciones posibles. Para obtener un conjunto de muestras, se han generado 100.000 muestras de tamaño 55 a partir de la muestra original, por una variante del procedimiento *Bootstrap*. La idea del *Bootstrap* es obtener, a partir de una muestra de tamaño n, muestras del mismo tamaño extraídas de la distribución empírica de la muestra original, o, lo que es lo mismo, mediante muestreo con reposición de tamaño n, de una población finita formada con los elementos de la muestra original. La ventaja fundamental de este procedimiento es que es un método no paramétrico.

La diferencia entre el procedimiento utilizado en este trabajo y el método *Bootstrap* estándar es que, en la muestra original hay una serie de elementos con información incompleta que se tratan asociándoles un valor que es función de los valores muestrales de los elementos con información completa. Siendo consecuentes con esta idea, al obtener las muestras sintéticas, el valor asociado a estos elementos no es fijo. Lo anterior es coherente con el principio de no utilizar en las estimaciones información no contenida en la muestra particular en la que se base la estimación. Dicho principio se violaría si se utilizara, como valor asociado a los elementos sin información completa, el obtenido para la estimación puntual con la muestra original. A partir de las 100.000 muestras anteriores y sus correspondientes estimaciones de $E(LDG_1)$, basadas en la media simple, media ponderada y mediana (con el tratamiento ya comentado de los elementos sin información completa), se han obtenido las correspondientes distribuciones (los escenarios son los mismos para todos los procedimientos de estimación).

GRÁFICO 8

DISTRIBUCIÓN DEL ESTIMADOR MEDIA



3.6.1. Intervalo para E(LGD_1) basado en la media simple

El gráfico 8 presenta la distribución empírica obtenida en las simulaciones y resume los resultados obtenidos.

La línea gris corresponde a la función de densidad de una variable aleatoria normal (36) con las mismas media y desviación típica que las de la distribución obtenida en el *Bootstrap*. Se observa que el percentil correspondiente al 5 % es LGD^p=9,7 %, y que el percentil correspondiente al 95 % es LGD^p=15,7 %. Por lo tanto, un intervalo de confianza al 90 % para la E(LGD) poblacional por el procedimiento *Bootstrap* viene dado por [9,7 %, 15,7 %]. En términos de la definición de incumplimiento DF_2, se transforma en [21,6 %, 35,2 %].

Se puede utilizar, alternativamente a la simulación, una aproximación analítica basada en mixturas de normales que produce resultados muy parecidos (37).

3.6.2. Intervalo para E(LGD_1) basado en la media ponderada

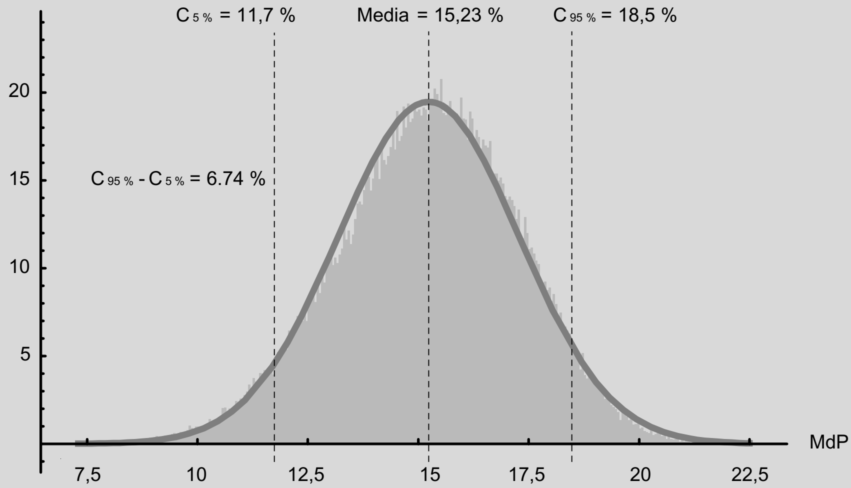
Utilizando la media ponderada por los importes, se obtiene una distribución con mayor dispersión. Además, la no normalidad de la distribución empírica es más acusada que en el caso anterior (gráfico 9).

(36) Se aprecia ligeramente la no normalidad de la media muestral, debida al tratamiento dado a los elementos sin información que hace perder la independencia de las observaciones.

(37) En efecto, dada la forma en que se tratan los elementos sin información, el proceso de generación de las muestras se puede descomponer en dos etapas. En la primera, se obtiene para cada muestra de tamaño n=55 el número de elementos r con información completa. Dada la muestra base, r sigue una distribución binomial de parámetros (n=55, p=40/55). Condicionado al valor de r obtenido, la distribución de la media muestral de la muestra sintética completa es, a partir de un valor de r suficientemente grande, aproximadamente normal con parámetros ($\mu=m_x$, $\sigma^2=s^2_x/r$). m_x y s^2_x son la media y la varianza de la muestra base sin los elementos con información incompleta. En el anejo 2 se detallan la función de distribución correspondiente y su comparación con la aproximación normal y los intervalos de confianza asociados.

GRÁFICO 9

DISTRIBUCIÓN DEL ESTIMADOR MEDIA PONDERADA



El percentil correspondiente al 5 % es LGD^p=11,7 %, y el percentil correspondiente al 95 % es LGD^p=18,5 %. Por lo tanto, un intervalo de confianza al 90 % para $E(LGD_1)$ poblacional por el procedimiento *Bootstrap* viene dado por [11,7 %, 18,5 %]. En términos de la definición de incumplimiento DF_2, se transforma en el intervalo [26,2 %, 41,2 %].

3.6.3. Intervalo para $E(LGD_1)$ basado en la mediana

El comportamiento de la distribución de las estimaciones de $E(LGD_1)$ basadas en la mediana de los valores observados de las pérdidas se resume en el gráfico 10.

Su comportamiento es totalmente diferente. Esto es debido a que la mediana, tal como se ha definido para el cálculo, es el punto medio del intervalo mediano (38), lo produce un efecto de acumulación sobre determinados valores. El percentil correspondiente al 5 % es LGD^p=7,95 %, y el percentil correspondiente al 95 % es LGD^p=13,31 %. Por lo tanto, un intervalo de confianza al 90 % para $E(LGD_1)$ poblacional por el procedimiento *Bootstrap* viene dado por [7,95 %, 13,31 %]. En términos de la definición de incumplimiento DF_2, se transforma en el intervalo [17,7 %, 29,7 %].

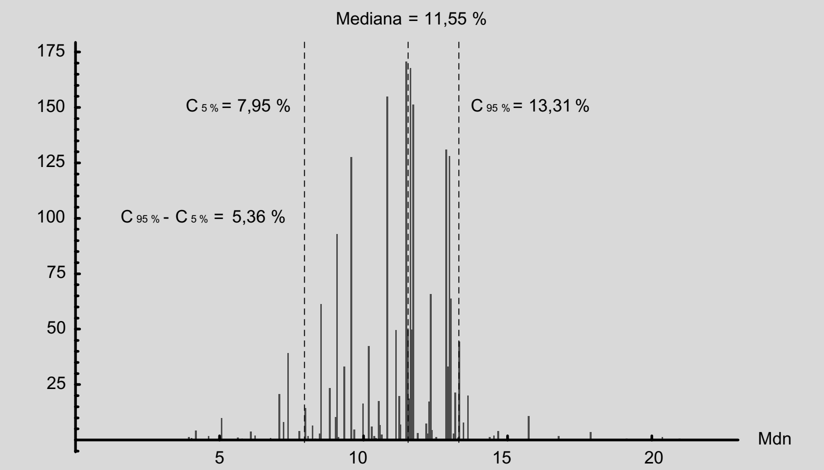
Una forma de evitar ese comportamiento tan irregular de la distribución del estimador es utilizar otra definición de mediana (39). Se va a cambiar ligeramente el procedimiento de estimación basado en la mediana.

(38) Nótese que la razón de que aparezcan intervalos medianos es que el muestreo a partir de la función de distribución empírica solo permite obtener observaciones iguales a las obtenidas en la muestra base (que son un número finito). Véase el anejo 3, sobre la función de distribución empírica.

(39) Otra forma es obtener las muestras de una función de distribución ligeramente diferente de la función de distribución empírica, permitiendo que se puedan alcanzar los valores intermedios.

GRÁFICO 10

DISTRIBUCIÓN DEL ESTIMADOR MEDIANA



En cada una de las muestras, en vez de calcular la mediana muestral de los elementos con información completa e imputarla como valor a los elementos sin información en esa muestra, se calculan los percentiles muestrales $C_{40\,\%}$ y $C_{60\,\%}$. A continuación, se toman como valores observados de los elementos sin información el resultado de una muestra uniforme en el intervalo ($C_{40\,\%}$, $C_{60\,\%}$) de tamaño igual al número de elementos sin información. Finalmente, se calcula la mediana de los valores así obtenidos. El gráfico 11 muestra el comportamiento de este estimador:

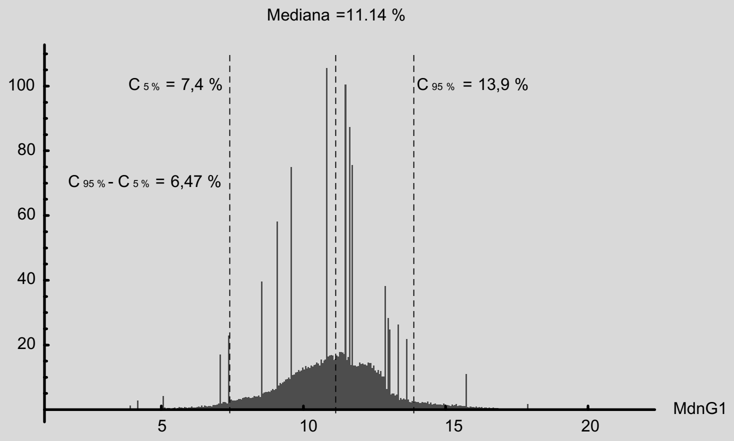
El comportamiento del estimador se suaviza parcialmente, pero, al seguir calculando la mediana para cada muestra, siguen apareciendo concentraciones en valores aislados. El percentil correspondiente al 5 % es LGD^p=7,4 %, y el percentil correspondiente al 95 % es LGD^p=13,9 %. Por lo tanto, un intervalo de confianza al 90 % para la E(LGD_1) poblacional por el procedimiento *Bootstrap* viene dado por [7,4 %, 13,9 %]. En términos de la definición de incumplimiento DF_2, se transforma en el intervalo [16,5 %, 31 %].

Para que desaparezca totalmente ese efecto, se va a modificar el procedimiento anterior. En este caso, la estimación para cada muestra no es la mediana. En su lugar, se repite el procedimiento explicado anteriormente para obtener los valores asociados a los elementos sin información. El gráfico 12 presenta la distribución muestral encontrada.

Se observa que desaparecen las concentraciones en valores individuales. El percentil correspondiente al 5 % es LGD^p=7,2 %, y el percentil correspondiente al 95 % es LGD^p=14,2 %. Por lo tanto, un intervalo de confianza al 90 % para la E(LGD_1) poblacional por el procedimiento *Bootstrap* viene dado por [7,2 %, 14,2 %]. En términos de la definición de incumplimiento DF_2, se transforma en el intervalo [16 %, 31,7 %].

GRÁFICO 11

DISTRIBUCIÓN DEL ESTIMADOR. MEDIANA G1



3.6.4. *Comparación entre los diferentes métodos de estimación*

El gráfico 13 muestra los resultados de la estimación de los intervalos al 90 % con cada uno de los métodos anteriores, ordenados de menor a mayor amplitud, asociados a la definición de incumplimiento DF_1.

El intervalo más estrecho es el asociado a la mediana (Mdn), seguido por el intervalo para el estimador basado en la media muestral (Md). Si se emplean medianas generalizadas (MdnGi) para suavizar el estimador basado en la mediana, los intervalos se agrandan y se colocan por detrás de

GRÁFICO 12

DISTRIBUCIÓN DEL ESTIMADOR. MEDIANA G2

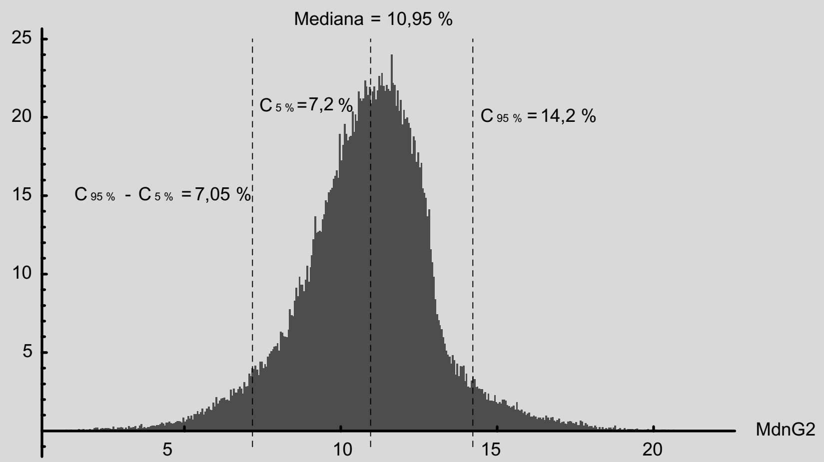
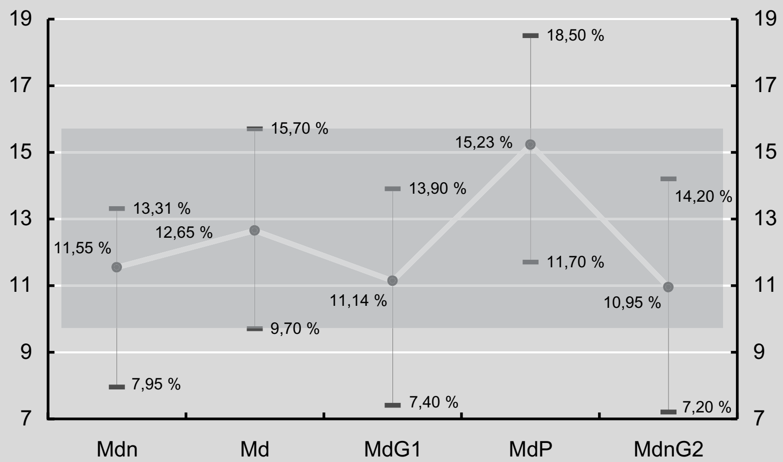


GRÁFICO 13

COMPARACIÓN DE LOS DIFERENTES ESTIMADORES



la media y de la media ponderada (MdP). El único intervalo que contiene todos los valores de los estimadores puntuales es el basado en la media.

4. INCORPORACIÓN DE INCERTIDUMBRE Y SESGOS EN LOS ELEMENTOS SIN INFORMACIÓN

Hasta ahora, no se ha tenido en cuenta explícitamente la incertidumbre existente en los elementos con información incompleta. La idea es asociar a dichos elementos una variable aleatoria (en vez de una función), cuyos parámetros dependen de los valores muestrales del resto de elementos con información completa.

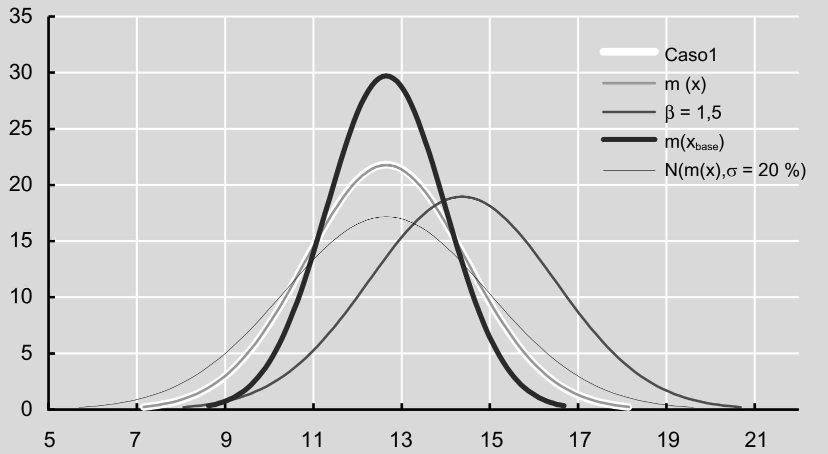
Supóngase que a cada elemento con información incompleta se le asocia como LGD una variable aleatoria $N[m(x), \sigma^2]$. El método es el mismo esquema *Bootstrap* descrito anteriormente, con la única modificación de que, en cada muestra sintética, la LGD de los elementos sin información completa se extrae de la distribución normal anterior, en la que su media m(x) depende del resto de elementos con información completa de la muestra.

En este caso, para la distribución basada en el estadístico media muestral, hay una aproximación analítica. En otros casos más complicados, en los que, por ejemplo, la varianza sea función de la muestra, $\sigma^2(x)$, se puede calcular la distribución eficientemente por procedimientos numéricos.

Si con la información disponible los elementos sin información se pueden clasificar en función de que su LGD (esperada) sea menor/mayor/igual que la del resto de elementos de la muestra, dicha información se puede incorporar al esquema anterior, introduciendo un sesgo respecto a la media del resto de los elementos de la muestra.

GRÁFICO 14

COMPARACIÓN DE DIFERENTES ESTIMADORES



El gráfico 14 compara los resultados obtenidos para el estimador basado en la media muestral, para diferentes esquemas de *Bootstrap*.

La curva $m(X_{base})$ representa la densidad cuando se asocia como LGD de los elementos con información incompleta la media obtenida en los elementos con información completa de la muestra base. Es la más concentrada de todas, pero el procedimiento no es coherente, ya que, dada una muestra sintética, para asociar un valor a los elementos sin información completa utiliza información no presente en dicha muestra. La densidad m(x) es la correspondiente a asociar a cada elemento sin información completa la media m(x) del resto de elementos con información completa presentes en la muestra sintética (Caso 1). En el gráfico 14, prácticamente coincide con la curva normal asociada al procedimiento clásico de ignorar los elementos con información incompleta. Cuando se considera explícitamente la incertidumbre asociada a los elementos con información incompleta, la dispersión alrededor de la media aumenta, como puede verse con el caso N[m(x), σ=20 %], en el que, en cada muestra sintética x, a los elementos con información incompleta se les asigna una realización de la distribución normal N[m(x), σ=20 %], donde m(x) es la media de los elementos con información completa de la muestra x. Por último, se presenta la densidad asociada a un sesgo de calidad en los elementos sin información, modelizado suponiendo que en cada muestra sintética los elementos sin información completa tienen asociado, como LGD, 1,5 veces la media muestral del resto de elementos de la muestra (β = 1,5).

5. COMPARACIÓN CON REFERENCIAS *(BENCHMARKS)* Y OTROS TRABAJOS SOBRE L.G.D.

En España, las entidades no suelen disponer de estimaciones de pérdida de sus operaciones hipotecarias basadas en el análisis detallado de los expedientes. Los valores utilizados en sus estimaciones de pérdida vienen, fundamentalmente, de referencias proporcionadas por consultores (en principio, basadas en su experiencia para operaciones similares en otros mercados) y/o en estimaciones de expertos (normalmente, de la propia entidad). Las cifras utilizadas por diferentes entidades para la severidad de las operaciones hipotecarias con particulares abarcan un rango que va desde el 20 % al 40 %, si bien a veces es difícil averiguar qué definición de incumplimiento hay detrás de esas estimaciones. Sin entrar en la definición de incumplimiento utilizada, es imposible efectuar una comparación homogénea, pero, si se supone que se utiliza una definición de incumplimiento similar a DF_2, ese intervalo se transforma, en términos de la definición DF_1 (impago durante 90 días), en el [8,96 %, 17,92 %]. Ese intervalo contiene el obtenido anteriormente con el nivel de confianza del 90 % para el estimador de E(LGD) basado en la media muestral, si bien el valor central, del 13,44 %, no es muy distinto del 12,65 % obtenido en este trabajo.

A primera vista, de entre los estudios disponibles, los resultados más fáciles de comparar con los de este trabajo parecen los derivados de estudios sobre severidad de titulizaciones de carteras hipotecarias. Un cálculo reciente de Standard&Poor´s (40) referido a «*Structured Finance Securities*» en Estados Unidos, estima las tasas de recuperación (100 – LGD) para «*residential mortgage-backed securities*» (RMBS), «*commercial mortgage-backed securities*» (CMBS), «*asset-backed securities*» (ABS) en el 61 %, 66 % y el 29 %, respectivamente. En términos de LGD (41), suponen el 39 %, 34 % y 61 %. Las diferencias entre las definiciones de incumplimiento utilizadas en estos estudios y en el presente trabajo imposibilitan la comparación homogénea.

Sobre el segmento de préstamos sindicados en Estados Unidos, Moody´s Investors Service ha publicado un trabajo (42) en el que se estima la tasa de recuperación para las categorías «*Senior Secure*» y «*Senior Unsecure*» en el 69,5 % y el 52,1 %, respectivamente (43). Expresado como LGD (44), supone unos porcentajes del 30,5 % y del 47,9 %.

Citibank, en un estudio (45) sobre pérdidas de préstamos comerciales e industriales en Latinoamérica, utiliza una definición de pérdida parecida a la utilizada en este trabajo y obtiene una E(LGD)=31,8 %. En este caso, la definición de incumplimiento utilizada está basada en categorías subjetivas: dudosos e incobrables *(doubtful, non accrual).*

---

(40) «Recoveries of defaulted U.S. Structured Finance Securities», Standard&Poors, 4 de septiembre de 2001.

(41) La definición de incumplimiento utilizada es el impago de intereses de los bonos en su fecha prevista o pérdidas en los préstamos subyacentes que hayan consumido la protección y causen pérdidas del principal de los títulos *(default rating D by S&P).*

(42) «Bank-Loan Loss Given Default», Moody´s Investors Service, noviembre 2000.

(43) Con una definición de pérdida basada en la cotización de la deuda un mes después del incumplimiento.

(44) La definición de incumplimiento utilizada incluye el impago (o retraso en el pago) de intereses o principal, la quiebra o suspensión de pagos y ciertas reinstrumentaciones.

(45) «Measuring Loss on Latin American Defaulted Bank Loans: A 27-Year Study of 27 Countries», Citibank, agosto 1998.

6. CONCLUSIONES

La metodología propuesta para la estimación de la severidad (LGD) permite:

- Incluir gastos judiciales y de adjudicación en el cálculo de la pérdida.
- El tratamiento específico de las adjudicaciones al propio grupo y de los elementos sin información.
- Facilitar la comparación con otras estimaciones de la severidad obtenidas a partir de diferentes definiciones de incumplimiento.
- Analizar la sensibilidad frente a variaciones en los parámetros.
- Contrastar la estabilidad ante variaciones en la muestra.

De la aplicación de esta metodología a una cartera de préstamos hipotecarios en España, concedidos a personas físicas, bien diversificada (geográficamente y por garantías), joven (operaciones con menos de seis años de vida), se obtiene:

- En la fecha de análisis, una estimación puntual de la LGD media basada en el impago, durante 90 días, del 12,65 %.
- Un intervalo de confianza al 90 % para dicha LGD media, dado por (9,7 %, 15,7 %).

Las estimaciones anteriores son razonablemente estables ante variaciones en la muestra. Por ejemplo, eliminando el 10 % de los elementos de la muestra de todas las formas posibles, el 90 % de las correspondientes estimaciones de la LGD media están comprendidas entre el 11,58 % y el 13,51 %.

Carteras de préstamos con, por ejemplo, gran variabilidad en el LTV de sus operaciones o vida media elevada precisan de una adecuada segmentación para calcular estimaciones de LGD (por segmento) estables en el tiempo.

ANEJOS

ANEJO 1: PÉRDIDA MEDIA EN UN INTERVALO DE TIEMPO

Se considera una operación cuya exposición en función del tiempo t viene dada por Exp(t). Se quiere encontrar E(L), la pérdida media de dicha operación en el intervalo de tiempo (0,1), descontada al origen, a partir de un modelo sencillo. Se supone conocida la PD, probabilidad de que dicha operación incumpla durante el intervalo considerado. Además, se supone que la distribución de probabilidad asociada al incumplimiento en el instante t condicionada a que la operación incumple en dicho intervalo, es una uniforme (0,1). Denominando E(LGD) la pérdida porcentual media en caso de incumplimiento, se tiene:

$$E(L) = (1-PD)*0 + PD*E_t\left(e^{-it}*Exp(t)*E(LGD)\right) = PD*E(LGD)*E_t\left(e^{-it}*Exp(t)\right) \quad [21]$$

Para el caso en el que no hay amortizaciones ni pagos de intereses durante el período, se tiene Exp(t) = Exp(0) * e^{rt}, donde r es el tipo de interés del préstamo, con lo que (suponiendo que r no es igual al tipo de descuento i) se llega a:

$$E(L) = PD*E(LGD)*Exp(0)*\int_0^1 e^{(r-i)t}dt = PD*E(LGD)*Exp(0)*\frac{e^{r-i}-1}{r-i} \quad [22]$$

En general, r–i es pequeño, con lo que, utilizando la aproximación de segundo orden $e^{r-i}-1 \cong r-i+\frac{(r-i)^2}{2}$, se obtiene:

$$E(L) \cong PD*E(LGD)*Exp(0)*(1+\frac{r-i}{2}) \quad [23]$$

Si i=r la fórmula es:

$$E(L) = PD*E(LGD)*Exp(0) \quad [24]$$

Para pasar a pérdidas medias de una cartera, E(LC), se pueden sumar las pérdidas medias de cada operación. Agrupando las operaciones de la cartera en k subcarteras con igual PD (igual clase de *rating)* e igual LGD:

$$E(L) = \sum_1^k \left(PD^k * E(LGD^k) * \sum_1^{jk} \left(Exp(0)_k * \frac{e^{r_{jk}-i}-1}{r_{jk}-i} \right) \right) \quad [25]$$

Usando la aproximación [23], se obtiene:

$$E(L) = \sum_1^k \left(PD^k * E(LGD^k) * \sum_1^{jk} \left(Exp(0)_k * (1+\frac{r_{jk}-i}{2}) \right) \right) \quad [26]$$

ANEJO 2: APROXIMACIONES ANALÍTICAS BASADAS EN MIXTURAS DE DISTRIBUCIONES

Se parte de una muestra base de tamaño n, que contiene j elementos con información completa (expedientes en los que se puede calcular directamente la pérdida) y n–j elementos con información incompleta. Se desea obtener las distribuciones de diferentes estimadores de la pérdida mediante un procedimiento *Bootstrap*. El algoritmo utilizado consiste en generar un número muy elevado de muestras sintéticas de tamaño n, mediante muestreo con reposición de la muestra base. Para calcular la estimación asociada a cada muestra así obtenida, hay que decidir el tratamiento de los elementos con información incompleta. En este trabajo se utiliza, como pérdida de dichos elementos, el valor de la estimación obtenida de los elementos con información completa.

Las distribuciones de los estadísticos *Bootstrap* utilizados (media simple, media ponderada y mediana) en el presente trabajo pueden representarse mediante sumas aleatorias de distribuciones. En efecto, el proceso de obtención de las muestras sintéticas de tamaño n en las simulaciones, a partir de la muestra base, se puede descomponer en dos etapas. En la primera se determina el número de elementos r con información completa presentes en la muestra de tamaño n que se va a generar. En la segunda se obtienen los r valores correspondientes a la pérdida asociada a las operaciones seleccionadas, mediante muestreo con reemplazamiento del conjunto de operaciones con información completa existentes en la muestra base.

La variable r de la primera etapa sigue una distribución binomial de parámetros n (número de elementos de la muestra base) y p (cociente entre el número de elementos con información y el número de elementos totales en la muestra base).

Denominando $H|_r$ la distribución condicionada del estadístico B que se está utilizando, se obtiene para la función de distribución incondicionada de B, que se denomina H(x), la expresión:

$$H(x) = k(n,p) \sum_{r=1}^{n} \binom{n}{r} p^r (1-p)^{n-r} H\big|_r (x) \qquad [27]$$

donde $k(n,p)=1/(1-(1-p)^n)$ es el inverso de la probabilidad de que r sea distinta de 0.

A.2.1. Caso particular del estimador basado en la media muestral, asociando a los elementos sin información la media muestral

En el caso particular del estimador basado en la media, la expresión para $H|_r$ es:

$$H\big|_r (x) = prob\left(\frac{1}{r} \sum_{i=1}^{r} L_i \le x \right) \qquad [28]$$

donde la distribución de los L_i es la distribución empírica asociada a los elementos con información completa de la muestra base. En este caso, con r suficientemente grande, se sabe que la distribución condicionada puede aproximarse a una normal:

$$H\big|_r(x) = prob\left(\frac{1}{r}\sum_{i=1}^{r} L_i \le x\right) \approx N\left(\mu = L^*,\ \sigma^2 = \frac{S^{2*}}{r}\right) \qquad [29]$$

donde L^* y S^{2*} son la media y la varianza de la distribución empírica asociada a los elementos de la muestra base con información completa. Para la mayor parte de los valores de p que pueden encontrarse en la práctica, el rango de valores relevantes de r (aquellos que aportan suficiente probabilidad en las fórmulas anteriores) hace posible utilizar la aproximación anterior para todos ellos. Con ello, finalmente, se obtiene una representación de la distribución del estadístico como una mixtura de normales.

$$H(x) = k(n,p)\sum_{r=1}^{n}\binom{n}{r} p^r (1-p)^{n-r}\, \Phi\left(\sqrt{r}\,\frac{x-L^*}{S^*}\right) \qquad [30]$$

Su función de densidad es:

$$h(x) = k(n,p)\sum_{r=1}^{n}\binom{n}{r} p^r (1-p)^{n-r}\, \frac{\sqrt{r}}{S}\, \phi\left(\sqrt{r}\,\frac{x-L^*}{S^*}\right) \qquad [31]$$

Los momentos respecto al origen se pueden calcular fácilmente a partir de la relación:

$$E\left[B^r\right] = k(n,p)\sum_{r=1}^{n}\binom{n}{r} p^r (1-p)^{n-r}\, E\big|_r\left[B^r\right] \qquad [32]$$

Aplicándolo para calcular la media y la varianza, se obtiene:

$$E[B] = L^*; \qquad V[B] = S^2\, k(n,p)\sum_{r=1}^{n}\binom{n}{r} p^r (1-p)^{n-r}\, \frac{1}{r} \qquad [33]$$

Desarrollando en serie de Taylor la función f(x)=1/x en el punto c=np, se sabe que para cada x en (a, b) existe un z perteneciente a ese intervalo, tal que:

$$f(x) = f(np) + f'(np)\,(x-np) + \frac{f''(z)}{2}\,(x-np)^2 = \frac{1}{np} - \frac{1}{(np)^2}\,(x-np) + \frac{1}{(z)^3}\,(x-np)^2 \qquad [34]$$

Si se consideran valores de x en el intervalo $\left(np - z_{\alpha}\sqrt{npq}, np + z_{\alpha}\sqrt{npq}\right)$, el término de error está acotado superiormente por su valor para $z = np - z_{\alpha}\sqrt{npq}$ y $x = np \pm z_{\alpha}\sqrt{npq}$. Con lo que se obtiene la cota para el término de error, dada por:

$$\frac{1}{(z)^3}(x-np)^2 \le \frac{1}{(np - z_{\alpha}\sqrt{npq})^3}\left(np + z_{\alpha}\sqrt{npq} - np\right)^2 = \frac{q\, z_{\alpha}^2 npq}{(np - z_{\alpha}\sqrt{npq})^3} \qquad [35]$$

Se observa que el término de error tiende a cero al crecer n. Eliminando ese término y sustituyendo en la ecuación [27], se obtiene:

$$V[B] \cong S^2\, k(n,p) \sum_{r=1}^{n} \binom{n}{r} p^r (1-p)^{n-r} \left(\frac{1}{np} - \frac{1}{(np)^2}(r-np)^2\right) \qquad [36]$$

Finalmente, haciendo operaciones, y teniendo en cuenta que k(n,p)$\rightarrow$ 1 cuando n es grande,

$$V[B] \cong \left(\frac{S^2}{np} - \frac{S^2}{(np)^2}\left(k(n,p)\, np - np\right)\right) \cong \frac{S^2}{np} \qquad [37]$$

Es decir, las mismas media y varianza que las asociadas al estimador media muestral, ignorando los elementos sin información completa.

A.2.2. Caso particular del estimador basado en la media muestral, asociando a los elementos sin información una variable aleatoria normal con varianza independiente de la muestra

Considérese el caso en el que los elementos sin información que aparecen en la muestra sintética se tratan asociándoles una variable aleatoria N(m(L), σ^2), donde L es la submuestra formada por los elementos con información completa y m(L) su media $m(L) = \frac{1}{r}\sum_{i=0}^{r} L_i$. La representación tación [27] sigue siendo válida, pero en este caso [28] toma la forma:

$$H\big|_r(x) = prob\left(\frac{1}{n}\left(\sum_{i=1}^{r} L_i + \sum_{i+1}^{n} x_j\right) \le x\right) = prob\left(\frac{1}{n}\left(r * m(L) + \sum_{i+1}^{n} x_j\right) \le x\right) \qquad [38]$$

donde cada x_j sigue una variable aleatoria con media m(L) y varianza fija σ^2. Haciendo operaciones, se llega a que la media muestral condicionada a r y a L:

$$\frac{1}{n}\left(r * m(L) + \sum_{i+1}^{n} x_j\right) \approx N\left(m(L), \frac{(n-r)}{n^2}\sigma^2\right) \qquad [39]$$

Pero se conoce la distribución de m(L) condicionado a r, $m(L)\big|_r \approx N(\mu = L^*, \frac{S^2}{r})$, donde L^* es la media muestral asociada a la muestra base. Con lo que, combinando ese resultado con [39], se obtiene, para la distribución de la media muestral condicionada a r:

$$\frac{1}{n}(r * m(L) + \sum_{i+1}^{n} x_j) \approx N(L^*, \frac{(n-r)}{n^2}\sigma^2 + \frac{S^2}{r}) \qquad [40]$$

Vemos que el efecto es que las distribuciones condicionadas tienen mayor dispersión en torno a la media. El resto es igual que en el caso anterior.

A.2.3. Caso particular del estimador basado en la media muestral, asociando a los elementos sin información la media muestral con sesgo

Supóngase que la LGD asociada a los elementos sin información completa es la media muestral del resto de los elementos de la muestra multiplicada por un coeficiente β. La media muestral condicionada a r tiene la forma:

$$\frac{1}{n}(r * m(L) + (n-r) * \beta * m(L)) = m(L) * (r + (n-r) * \beta) \approx N(L^*(r + (n-r) * \beta), \frac{S^2}{r}(r + (n-r) * \beta)^2) \qquad [41]$$

En este caso, las distribuciones condicionadas tienen diferente media y varianza, siempre que β>1. El resto es igual que en el caso anterior.

Análogamente, para la media ponderada y la mediana se pueden obtener representaciones similares.

ANEJO 3: FUNCIÓN DE DISTRIBUCIÓN EMPÍRICA

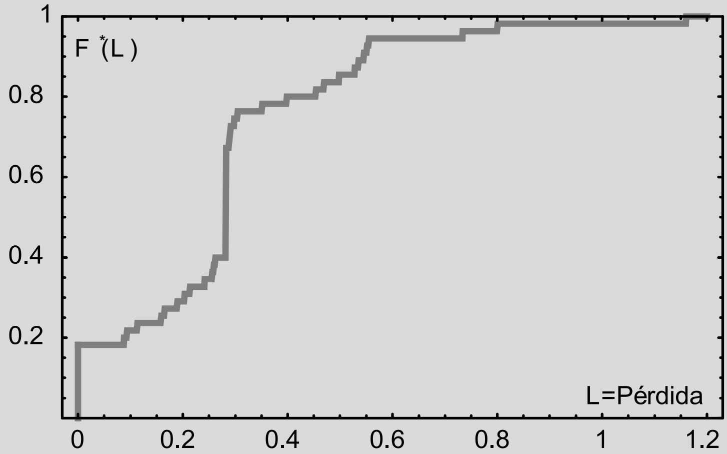
Dadas n observaciones de una variable aleatoria X, se denomina función de distribución empírica F*(x) a la función que asocia a cada x el cociente entre el número de elementos de la muestra menores o iguales que x y n.

La representación de la función de distribución empírica para la pérdida obtenida en la muestra de expedientes analizados es la que recoge el gráfico A.3.1.

El salto inicial correspondiente al valor L=0 se corresponde con los expedientes sin pérdida. El salto existente en el valor 0,28 corresponde al tratamiento dado a los elementos sin información, a los cuales se les asigna el promedio del resto de valores.

GRÁFICO A.3.1

FUNCIÓN DE DISTRIBUCIÓN EMPÍRICA F*

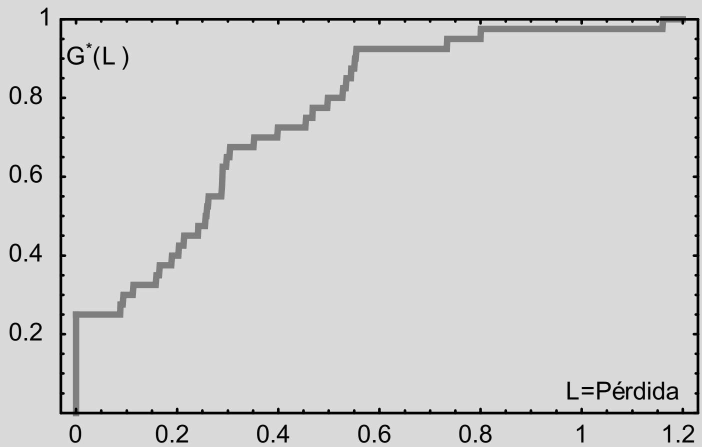


Para generar los procesos de *Bootstrap,* se ha utilizado la función de distribución empírica G* asociada a los elementos con información suficiente, que se presenta en el gráfico A.3.2.

En una primera etapa, se obtiene para cada muestra el número de elementos con información completa que la componen, mediante muestreo de una binomial. Para estos elementos, utilizando la función de distribución empírica anterior G*, se obtienen sus pérdidas. Se calcula el estadístico que se va a utilizar para la estimación (media simple, media ponderada o mediana) de los valores anteriores, y finalmente se asocia a los elementos con información incompleta dicho valor.

GRÁFICO A.3.2

FUNCIÓN DE DISTRIBUCIÓN EMPÍRICA G*



ANEJO 4: TIPOS DEL IMPUESTO SOBRE TRANSMISIONES PATRIMONIALES

A continuación se exponen los distintos tipos del Impuesto sobre Transmisiones Patrimoniales (ITP) afecto a bienes inmuebles, según Comunidades Autónomas (en vigor a diciembre de 2001).

Impuesto sobre Transmisiones Patrimoniales afecto a bienes inmuebles

Comunidades Autónomas	*Porcentaje aplicable*
Murcia, Valencia	4 %
Resto de Comunidades	6 %
Aragón, Baleares, Cataluña, Galicia, Madrid	7 %

BIBLIOGRAFÍA

BANCO DE ESPAÑA. *Circulares 4/1991, 9/1999* y *4/2000*.

BASEL COMMITTEE ON BANKING SUPERVISION (2001). *The New Basel Capital Accord. Consultative Document.*

BETANCOURT, L. (1999). «Using Markov Chains to Estimate Losses from a Portfolio of Mortgages», *Review of Quantitative Finance and Accounting,* mayo, pp. 303-317.

CARTY, L.V. y LIEBERMAN, D. (1996). *Defaulteded Bank Loans Recoveries,* Moody´s Inverstors Service, noviembre.

DEGROOT, M. H. (1988). *Probabilidad y estadística,* Addison-Wesley Iberoamericana.

DIACONIS, P. y EFRON B. (1983). *Computer-intensive methods in statistics,* Scientific American, mayo, pp. 116-130.

EFRON B. y TIBSHIRANI, R. J. (1993). *An introduction to the bootstrap,* Nueva York, Chapman & Hall.

FELLER, W. (1996). *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley & Sons, Inc.

GUPTON, G. M., GATES D. y CARTY, L. V. (2000). *Bank-Loan Loss Given Default,* Moody´s Inverstors Service, noviembre.

HU, POLLSEN, THOMPSON y HEDMAN (2001). *Recoveries of Defaulted U.S. Structured Finance Securities,* Standard&Poors, septiembre.

HURT, L. y FELSOVALYI, A. (1998). *Measuring Loss on Latin American Defaulted Bank Loans: A 27-year study of 27 countries,* Citibank, Nueva York, agosto.

MORAL, G. y OROZ, M. (2002). *Interest Rates and LGD estimates*.