POLITICAS ÓPTIMAS DE GESTIÓN DE ACTIVOS MEDIANTE PROGRAMACIÓN DINÁMICA

Pablo Bru Ribera

Trabajo de investigación 013/005

Master en Banca y Finanzas Cuantitativas

Tutores: Dr. Francisco Álvarez
Dr. Alfredo García-Hiernaux

Universidad Complutense de Madrid

Universidad del País Vasco

Universidad de Valencia

Universidad de Castilla-La Mancha

www.finanzascuantitativas.com

Políticas óptimas en la gestión de activos mediante optimización dinámica. *

Pablo Bru Ribera**

QFB, Universidad Complutense de Madrid, Spain

Tutores : Francisco Álvarez y Alfredo García-Hiernaux.

13 de junio de 2013

Resumen

En este trabajo nos centramos en resolver el *trade-off* entre la mejor política de compras de un conjunto de activos financieros y el aprendizaje que se genera en el proceso en un espacio temporal limitado. Tenemos en cuenta no sólo los posibles pagos futuros sino también la aversión al riesgo del comprador, así como el posible aprendizaje generado y la minimización del riesgo. Usando optimización dinámica con aprendizaje Bayesiano, presentamos un método de solución exacto que maximiza la utilidad esperada con aversión al riesgo y con horizonte temporal finito. Como resultado obtenemos la política óptima: un conjunto de órdenes de compra secuenciadas en el tiempo que maximizan la utilidad esperada.

*Agradecimientos.. XXX

**e-mail: bruribera@gmail.com Corresponding author.

Índice

1. Introducción

En la naturaleza de los problemas dinámicos bajo incertidumbre está la interacción entre el conjunto de decisiones que pueden tomarse y el aprendizaje que estas aportan. Las decisiones tomadas determinan no solo los posibles pagos futuros sino también el conocimiento futuro que generan y cómo condiciona este a las futuras decisiones. En numerosas ocasiones, los posibles rendimientos y el futuro aprendizaje entran en conflicto. Este *trade-off* surge en muchos problemas económicos y ha motivado numerosos trabajos en muy diversas ramas [3].

En este trabajo nos centramos en esta interacción en un espacio temporal limitado. Usando optimización dinámica con aprendizaje Bayesiano, presentamos un método de solución exacto maximizando la utilidad esperada con aversión al riesgo y con horizonte temporal finito. En esta línea, Álvarez y García-Hiernaux [2] están realizando trabajos para solucionar problemas relacionados con la obtención de la política óptima en la secuenciación de anuncios. El presente trabajo adapta el método de solución al contexto de las inversiones financieras, en donde las posibles pérdidas juegan un papel adicional que no está presente a la hora de hacer anuncios. Nos situamos, pues, en el lugar de un comprador que quiere maximizar la utilidad de sus inversiones. A diferencia de lo que se hace en el análisis media-varianza tradicional en el cual la minimización del riesgo se lleva a cabo en un único periodo, nosotros proponemos una secuenciación de las compras en multiples periodos que maximicen la utilidad total [4]. Un elemento crucial en el problema es la finitud del horizonte temporal que, combinado con la finitud en el número de activos que pueden comprarse, limita la capacidad del aprendizaje.

En el apartado 2 se describe el modelo de forma detallada. Comenzamos planteando el problema de un inversor que se enfrenta a la decisión de cómo gestionar la compra de una cantidad finita de activos de un mismo tipo. El inversor debe decidir la forma óptima de cómo secuenciar las compras maximizando una determinada función de utilidad. Las compras que realiza tienen una determinada probabilidad de éxito que se desconoce. Las creencias iniciales sobre la probabilidad de éxito se asume que se distribuyen con una distribución *Beta*, y se actualizan como consecuencia de las ventas/vencimientos de los activos observados en el periodo inmediato anterior. En las subsecciones siguientes se detalla la función de utilidad y el aprendizaje Bayesiano, así como el método de solución.

Como mencionamos anteriormente, presentamos el problema del inversor como un problema de optimización dinámica, y determinamos la política óptima de compras numéricamente mediante un algoritmo de programación dinámica. El apartado 3 muestra un análisis numérico de las políticas óptimas. En él se presentan simulaciones mediante Monte-Carlos de los posibles valores que puede tomar la política óptima en función de los estados de la naturaleza factibles. Finalmente se realiza un análisis de sensibilidad de los parámetros determinantes en el problema.

2. El Modelo

2.1. Planteamiento del problema

Supongamos que un banco de inversiones delega en un *trader* la inversión en activos financieros. Para cubrirse de eventuales pérdidas limita el presupuesto de negociación a 40.000 €, el cual le permite comprar 40 activos. Por lo tanto, el *trader* se planteará cuál debería ser su política óptima de compras.

En un primer momento, en la cabeza del *trader* se conjugan tres ideas fundamentales en la elección de la política óptima (aunque no las únicas). Estas son: posibles ganancias, posibles pérdidas y aversión al riesgo. La aversión al riesgo es la actitud que muestra el *trader* al enfrentarse a un juego en el que puede ganar o perder con una cierta probabilidad. Lo que se busca es resolver el *trade-off* que el problema plantea: minimizar los riesgos y maximizar los beneficios. Pero aún deberá tener en cuenta otros parámetros que determinan la elección.

En la compra de cualquier activo financiero se dispone de unas determinadas creencias iniciales acerca de cuál será la tendencia del activo subyacente. Esta información o creencias iniciales influirán en la compra o no y qué cantidad. Se tiene, por tanto, una serie de parámetros a tener en cuenta: cantidad incial de activos que podemos comprar, $x(0)$; posibles pérdidas y ganancias, W y L (*Winnings and Losses*); parámetro de aversión al riesgo, ρ ; y unas creencias inciales reflejadas a través de $(\alpha(0), \beta(0))$.

$$(x(0), W, L, \alpha(0), \beta(0), \rho) \tag{1}$$

Una vez establecidos los parámetros que condicionan la elección, el objetivo del *trader* es encontrar la mejor política de compras que puede llevar a cabo. La política óptima en este caso será un conjunto de órdenes de compra secuenciadas en el tiempo que determinan la cantidad de activos a comprar. Entendemos por compra de un activo financiero la adquisición del mismo al comienzo de un periodo. Cuando el periodo finaliza se produce su venta o vencimiento (en función del tipo de activo) e inmediatamente se produce la siguiente compra al comienzo del siguiente periodo.

El modelo está sometido a ciertos supuestos sobre sus parámetros.

1. El presupuesto es exógeno: el presupuesto condiciona la cantidad $x(0)$ máxima de activos disponible;

2. en el mercado existe una cierta probabilidad δ de que las operaciones tengan éxito. Esta probabilidad es contante en el tiempo y es, a priori, desconocida;

3. la información se recoge y actualiza simultanemente al final de cada periodo de tiempo;

4. las pérdidas y ganancias (L y W) son conocidas y constantes;
5. la función de utilidad que evalúa el *riesgo-beneficio* es una función de utilidad *CARA Bernouilli*.

En esta línea, suponemos que el tiempo es discreto y el horizonte de ejecución del algoritmo es finito por construcción[1], $t = 0, 1, 2..., T$. Así pues, el algoritmo, ejecutado por el *trader*, proporciona la cantidad óptima de activos que deberían ser comprados en cada periodo de tiempo t. A esta cantidad óptima de activos la llamaremos $u(t)$. Notar que esta cantidad $u(t)$ se determina dadas las creencias iniciales en $t = 0$, es decir, que el algoritmo es capaz de analizar todos los posibles estados futuros y determinar qué camino es el que maximiza una determinada función de utilidad (es decir, es óptimo).

2.2. Función de Utilidad Esperada y aprendizaje

Dijimos que si tenemos una cantidad de dinero tal que nos permite comprar un total de $x(0)$ activos, habría una cantidad $u(t)$ que es la cantidad óptima de compras en t. Digamos que $y(t)$ es una variable aleatoria que indica el número de activos $u(t)$ que han reportado beneficios dada la probabilidad de éxito δ. En otras palabras, $y(t)$ es el número de operaciones que han tenido éxito en t y se ditribuye como una $Binomial(\delta, u(t))$.

La compra de activos $u(t)$ óptima se realiza al comienzo de cada periodo de tiempo. Al final del periodo, cuando el activo es vendido o llega a vencimiento (dependiendo del tipo de activo), se observa la realización de la variable aleatoria $y(t)$ e inmediatamente después, cuando comienza el siguiente periodo, se actualizan creencias y se vuelve a determinar $u(t + 1)$.

El número de operaciones que tienen éxito $y(t)$ y las que no, $u(t) - y(t)$, ponderadas por sus correspondientes pérdidas y ganancias, nos proporciona los pagos monetarios que se obtiene en cada operación. Sea $r(t)$ estos pagos:

$$r(t) = y(t)W + (u(t) - y(t))L, \qquad 0 \leq y(t) \leq u(t) \tag{2}$$

donde las pérdidas han sido definidas negativas: $L < 0$.

Suponemos que el comprador evalúa su utilidad esperada mediante la función ex-ante de Von Neumann-Morgenstern, usando una función de utilidad CARA Bernouilli [2]. Elegimos este tipo de función porque es de fácil tratabilidad matemática, supone que el individuo es racional en su elección y nos proporciona información ordinal acerca de sus preferencias de elección. Las funciones CARA itroducen independencia del efecto

[1] El algoritmo de resolución de la política óptima tiene mecanismos para evitar la entrada en bucles o políticas que derivan en no ejecutar ninguna orden, lo que evita que el horizonte temporal sea infinito. Esto incurre en una intromisión desde el exterior en lo que sería la política óptima

riqueza, es decir, para el *trader* de nuestro ejemplo la riqueza acumulada o perdida en el pasado no condiciona su elección en el presente. En este sentido, los incrementos en su utilidad al incrementar en una unidad monetaria su riqueza es siempre la misma.
La utilidad esperada del *trader* es:

$$v\left(u\left(t\right), y(t)\right) := E\{-e^{-\rho r(t)} \mid u\left(t\right)\} \tag{3}$$

donde ρ es el parámetro (constante) de aversión absoluta al riesgo (ver Apéndice A.1).

Como vemos, la utilidad esperada en (3) es condicional a las creencias que tenga en t (a través de la variable $y(t)$) según sea la probabilidad de éxito. O sea, que depende en última instancia de δ, la verdadera probabilidad de éxito. La relación entre la verdadera probabilidad δ y las creencias iniciales $(\alpha(0), \beta(0))$ es que la verdadera probabilidad se modeliza a través de una $Beta(\alpha(0), \beta(0))$. Estas creencias se actualizan de forma Bayesiana según las expresiones:

$$\alpha(t+1) = \alpha(t) + y(t) \qquad \beta(t+1) = \beta(t) + u(t) - y(t) \tag{4}$$

y las creencias posteriores también se distribuyen como una *Beta*. Esta actualización de los parámetros es importante pues tiene implicaciones en la política óptima. Por ejemplo, si se decide comprar una cantidad $u(t)$ y tenemos el máximo número de éxitos posibles, es decir, $y(t) = u(t)$, como consecuencia tendremos para el siguiente periodo $t+1$ (ver ecuación (4)) que: $\alpha(t+1) > \alpha(t)$ y $\beta(t+1) = \beta(t)$. Un aumento en α con respecto a β supone que la distribución *Beta* de las creencias desplaza su valor esperado hacia valores más próximos a 1 (mayor probabilidad de éxito), con lo que el algoritmo "aprende" cuando observa las realizaciones.

Por otro lado, la dinámica del número de compras que restan por hacer, $x(t)$, viene dada por:

$$x(t+1) = x(t) - u(t) \tag{5}$$

con la restricción:

$$x(t+1) \geq 0 \qquad \forall t \in \{0, 1, ...\} \tag{6}$$

El número de éxitos acumulados hasta el momento t viene dado por $\gamma(t)$:

$$\gamma(t) := \sum_{s=0}^{t-1} y(s) \tag{7}$$

y su dinámica de actualización es:

$$\gamma(t+1) = \gamma(t) + y(t) \tag{8}$$

Finalmente, introduzcamos el factor de descuento λ. La solución del problema de programación dinámica busca la secuencia $\{u(t)\}_{t=0}^{T}$ que maximiza la suma descontada de los ingresos esperados:

$$\sum_{t=0}^{T} \lambda^t v(u(t), \gamma(t)) \tag{9}$$

siendo T el tiempo final para el cual $x(T) = 0$.

2.3. Método de solución: Optimización Dinámica

En esta sección se presenta el método de optimización del algoritmo basado en el principio de optimalidad de Bellman. El modelo más simple de optimización dinámica basado en la ecuación de Bellman tiene dos aspectos fundamentales: 1) un sistema dinámico en tiempo discreto, y 2) una función valor $J(.)$ que es aditiva en el tiempo. El algoritmo es llamado dinámico porque incorpora algunas variables que evolucionan en el tiempo bajo la influencia de las decisiones tomadas [1]. Por ejemplo, el estado:

$$x_{t+1} = f_t(x_t, u_t, \gamma_t), \qquad t = 0, 1, ..., T$$

donde x_t representa el estado del sistema en t y recoge la información que es relevante hasta el momento; u_t es la llamada variable de control y surge del proceso optimizador en $t-1$; γ_t es un parametro aleatorio y f_t es la función que describe la forma en que se actualizan los parámetros del sistema.

La política óptima es una aplicación que va desde el espacio de la *función valor* $J(x, u, \gamma)$ (resultado de la ecuación de Bellman) hasta el espacio de las variables de control $u(t)$. La terna (x, u, γ) determina la política óptima. En relación al problema de gestión de activos al que nos enfrentamos, lo que realmente nos interesa es el valor $u(t)$ (cantidad óptima de activos que comprar), y la ecuación de Bellman nos proporciona el valor $J(x, u, \gamma)$, necesario para encontrar el $u(t)$ óptimo. Antes de presentar la ecuación de Bellman de forma explícita veamos las expresiones del resto de variables que la componen y cómo éstas se actualizan en el tiempo.

2.3.1. La Ecuación de Bellman

La Programación Dinámica se basa en el *Principio de Optimalidad de Bellman*, principio por el cual, sea $\pi^* = \{\mu_0^*, \mu_1^*, ..., \mu_N^*\}$ la política óptima para un problema dado, cualquier estado intermedio del problema, por ejemplo el estado x_i, tiene como política óptima la serie truncada $\{\mu_i^*, \mu_{i+1}^*, ..., \mu_N^*\}$ del problema original [1]. Dicho con otras palabras, los subproblemas del problema también son soluciones óptimas.
Realmente, el *Principio de Optimalidad* no es más que la división de un problema grande en subproblemas más pequeños: primero resolviendo la política óptima para el último paso (en nuestro caso cuando $x(T-1) = 1$); después resolviendo la política óptima para los dos últimos pasos, y así sucesivamente hasta resolver el problema en $t = 0$.

En la sección 2.3 se presenta la función valor como aquel valor determinado por la terna (x, u, γ) que nos da la política óptima. En concreto, $J(x, u, \gamma)$ es la utilidad esperada descontada bajo la política óptima de cualquier subproblema dado por (x, u, γ):

$$J(x, \boldsymbol{u}, \gamma) = \max_{\boldsymbol{u} \in \Omega(x)} \{v(\boldsymbol{u}, \boldsymbol{\gamma}) + \lambda \sum_{\hat{\gamma} \in \theta(\boldsymbol{u}, \boldsymbol{\gamma})} J(x - \boldsymbol{u}, \boldsymbol{u}, \hat{\boldsymbol{\gamma}}) \Pr(\hat{\boldsymbol{\gamma}} \mid \boldsymbol{u}, \boldsymbol{\gamma})\} \tag{10}$$

donde $\Omega(x) = \{u\} : x - u \geq 0$, es el conjunto de todas las variables de control posibles, y $\theta(\boldsymbol{u}, \boldsymbol{\gamma}) = \{\hat{\gamma}|u, \gamma\} : \gamma \leq \hat{\gamma} \leq \gamma + u$, es el conjunto de todos los posibles éxitos dados (u, γ).

3. Análisis numérico

En esta sección se analizan los valores numéricos de las variables representativas del problema, así como su influencia en la determinación de la política óptima.
En primer lugar se introduce un ejemplo sencillo de política óptima. A continuación se hará un análsis de sensibilidad. En él se variará el parámetro de aversión al riesgo ρ y veremos la influencia en el horizonte temporal. Más tarde, haremos *mean-preserving spread*, que consiste en variar los parámetros W y L manteniendo constante el valor esperado de la inversión mientras varía su varianza y veremos las implicaciones que esto tiene en la política óptima de compras. Finalmente, realizamos un estudio de la influencia de las creencias inciales en la política óptima del inversor.

3.1. Un ejemplo sencillo de política óptima

Es importante entender la metodología del algoritmo en el contexto de un número incial reducido de inversiones. Para ello realizamos un ejemplo sencillo con $x(0) = 4$. La Tabla 1 es un ejemplo[2]. Como vemos, $x(0) = 4$ corresponde al instante inicial $t = 0$. El algoritmo, teniendo en cuenta la función de utilidad esperada y mediante la ecuación de Bellman, induce que lo óptimo es invertir comprando $u = 1$ activos.

Puede resultar curioso el hecho de secuenciar las compras, más aún cuando tenemos un factor de descuento. La respuesta es muy lógica: el *trader* es averso al riesgo, por lo que no está sólo valorando los posibles flujos futuros sino que está teniendo en cuenta el riesgo asociado ex-ante mediante la función de utilidad esperada de Von Neumann-Morgenstern (3). En la práctica se traduce en que "prefiere" posponer inversiones para periodos futuros y reducir así el riesgo global [5]. El ejemplo más sencillo sería aquel en que tenemos dos compras por hacer y las secuenciamos. Secuenciar las compras reduce el riesgo, pues cuando realizo la segunda compra sé el resultado de la primera. A su vez, la segunda compra se verá afectada por el factor de descuento. Ambos efectos, la reducción del riesgo y los flujos descontados, son evaludods por la política óptima.

[2] En la Tabla 1 se han omitido las filas con $x = 0$ porque son triviales. Evidentemente en ellas $u = 0$ y $J = 0$

Tabla 1: Ejemplo de política óptima. Valores de los parámetros: $x(0) = 4$, $\alpha(0) = 1$, $\beta(0) = 1$, $W = 10$, $L = 10$, $\rho = 0{,}007$, $\lambda = 0{,}9$

x	γ	u	J
1	0	0	0
1	1	0	0
1	2	1	1.50
1	3	1	5.50
2	0	0	0
2	1	1	0.18
2	2	2	8.60
3	0	0	0
3	1	2	8.26
4	0	1	3.22

¿Qué ocurre cuando posponemos inversiones? Posponer las inversiones supone dejar jugar a la verdadera naturaleza del problema. En este punto juega la verdadera probabilidad de éxito δ. Cuando pasamos de un periodo al siguiente observamos el número de extios obtenidos en el proceso. Recordemos la ecuación (4). Las realizaciones de los posibles estados de la naturaleza del problema determinan los futuros valores de α y β, esto es, las creencias sobre la verdadera probabilidad (recuerda que las creencias sobre la probabilidad de éxito δ la modelizamos mediante una distribución $Beta(\alpha(0), \beta(0))$). Aunque δ es desconocida por el inversor, en el proceso de inversión el algoritmo va aprendiendo sobre ella.

En esta línea, hemos realizado 100 simulaciones por Monte-Carlo de las posibles trayectorias que tomaría u fijado un valor para δ (ver Figura 1). Las realizaciones de la variable aleatoria $y \sim Binomial(\delta, u)$ (número de éxitos) es la que hace diferir unos caminos de otros. Como vemos en la Figura 1 el u de partida es el mismo para todos porque todos comparten los valores inciales y, por tanto, la política de partida es única. Un punto fuerte a tener en cuenta es que, sea cual sea el camino determinado por la naturaleza aleatoria del problema, todas las decisiones tomadas son óptimas. A esta configuración de los paramámetros la llamamos benchmark.

Por último, destacar una trayectoria en particular trazada en *negrita* que destaca del resto. Esta trayectoria ha sido trazada como si δ fuera conocida. Es decir, que el número de aceptaciones es el valor promedio $u\delta$ (esperanza de y, ver Apéndice A.2). A esta trayectoria la llamamos *camino promedio*. Esta trayectoria es representativa de la verdadera situación del mercado, que es deconocida. Observamos que las simulaciones se ajustan solo parcialmente. Esto se debe a que realmente δ es deconocida y al cominenzo del proceso de aprendizaje el algoritmo explora la posibilidad de comprar más (por

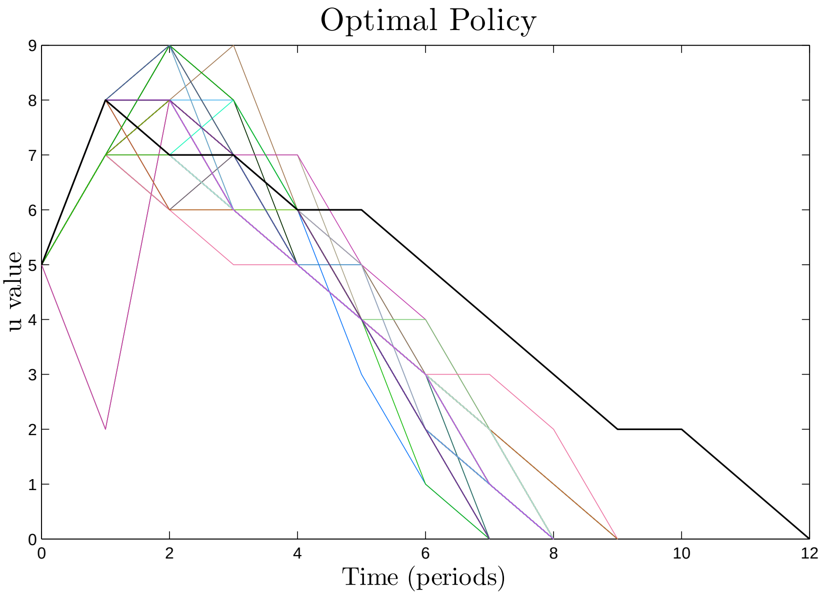

Figura 1: Posibles caminos de u. Valores inciales (benchmark): $x(0) = 40$, $\alpha(0) = 1$, $\beta(0) = 1$, $W = 10$, $L = 5$, $\rho = 0{,}01$, $\lambda = 0{,}9$, $\delta = 0{,}7$. Número de iteraciones de Monte-Carlo: 100.

ello los máximos valores de u pertenecen a simulaciones). Mayores valores de u supone terminar antes en el tiempo. Una vez explorada la posibilidad de comprar más que el *camino promedio* y ser refutada, el promedio de las simulaciones se sitúa por debajo del *camino promedio*.

3.2. Análisis de sensibilidad

En el análisis de sensibilidad tratamos de visualizar el efecto que tienen los parámetros sobre las políticas óptimas. Nos sirve como referencia el caso benchmark (Figura 1). A partir de él modificamos sus parámetros para ver la influencia en la política óptima del inversor.

3.2.1. Influencia de ρ

El parámetro de aversión absoluta al riesgo juega un papel central en la detrminación de la política óptima. El algoritmo quiere optimizar riesgo y beneficios. Cuando ρ es alto la política se vuelve conservadora y prefiere posponer las compras a la espera de observar los resultados y "aprender". Al final, posponer compras es reducir el riesgo. Como consecuencia la política se extiende en el tiempo. Las Figuras 2 y 3 ilustran este hecho.

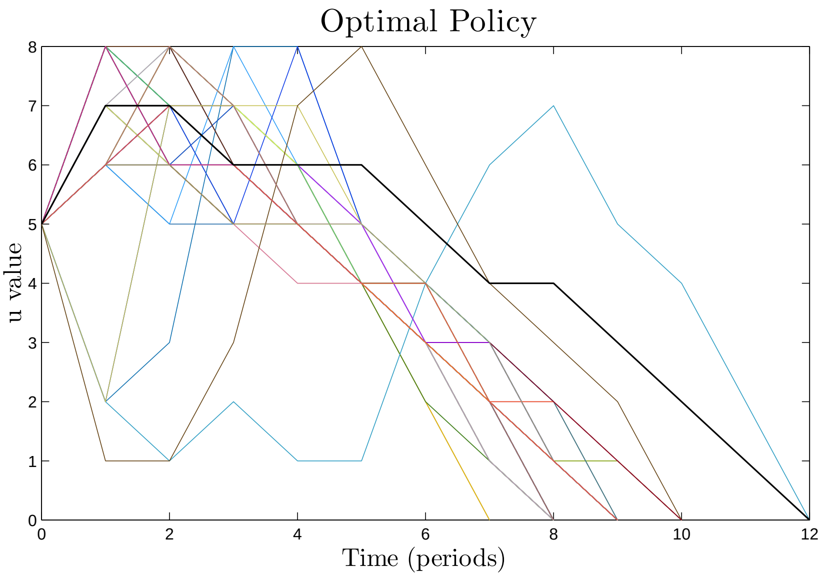

Figura 2: Posibles caminos de u. Influencia del parámetro ρ. Valores inciales: $x(0) = 40$, $\alpha(0) = 1$, $\beta(0) = 1$, $W = 10$, $L = 5$, $\boldsymbol{\rho = 0{,}012}$, $\lambda = 0{,}9$, $\delta = 0{,}7$. Número de iteraciones de Monte-Carlo: 100.

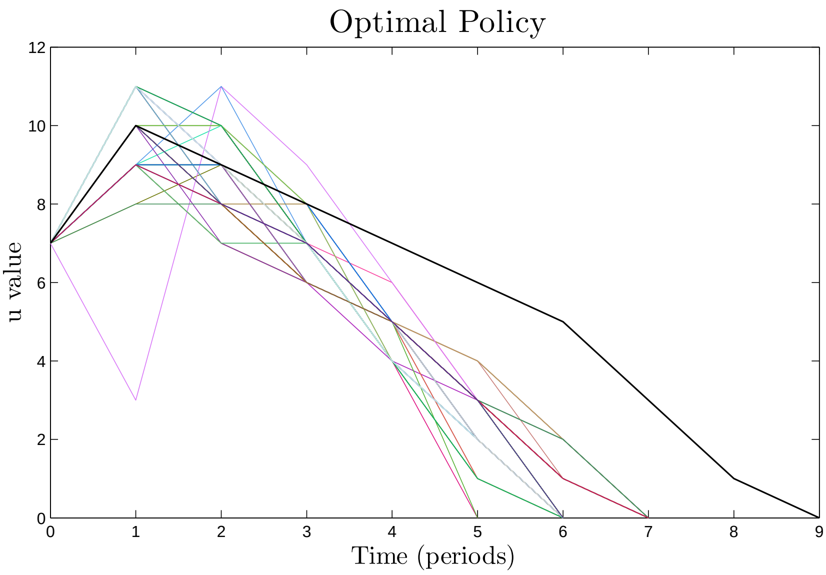

Figura 3: Posibles caminos de u. Influencia del parámetro ρ. Valores inciales: $x(0) = 40$, $\alpha(0) = 1$, $\beta(0) = 1$, $W = 10$, $L = 5$, $\boldsymbol{\rho = 0{,}007}$, $\lambda = 0{,}9$, $\delta = 0{,}7$. Número de iteraciones de Monte-Carlo: 100.

Además, cuando la aversión al riesgo es alta (baja) $u(0)$ es menor (mayor) que cuando es baja (alta). Consecuencia previsible desde un punto de vista intuitivo y quiere decir "no me importa empezar comprando más porque mi aversión al riesgo es baja".

Otra consecuencia visible es que cuando ρ es alto, Figura 2, la política es mucho más variable. Es decir, vemos más caminos que comienzan reduciendo el número de órdenes de compra[3]. Esto se debe a que, pese a que estadísitcamente el número de éxitos es proporcinoal, cuando una operación sale mal el algoritmo incurre en que está asumiendo demasiado riesgo y retrae su política. Como en este caso el mercado está en, digamos, una situación favorable debido a que δ es alto, los caminos que empiezan a la baja tienden a recuperarse. Es importante tener en cuenta estos sucesos menos probables porque pueden suceder y, aún así, la situación se resuelve de forma óptima para el *trader* dadas sus creencias iniciales y su aversión al riesgo.

Si hacemos ρ suficientemente grande, el algoritmo resolverá que la política óptima es comenzar con $u(0) = 0$ (ver nota al pie 1) lo que equivale a no realizar ninguna compra. Esto es lógico desde el punto de vista del *trader* ya que si su aversión al riesgo es demasiado alta realmente prefiere no invertir y no incurrir en ningún riesgo.

3.2.2. Mean-preserving spread

Esta sección analiza una situación importante en el contexto de las finanzas: dada una inversión con un valor esperado en sus rendimientos y un riesgo asociado, ¿cómo afectan las variaciones en los flujos esperados en las decisiones del *trader*? Suponemos, además, que las variaciones en los flujos son tales que el valor esperado de la inversión es constante y lo que cambia es la varianza.

Es decir, queremos observar las consecuencias en la política óptima cuando:

$$E(dr \mid u) = (dW - dL)E(y \mid u) + udL = 0 \tag{11}$$

La ec. (11) conduce a la relación (ver demostraciones en el Apéndice B.1):

$$-\frac{\triangle W}{\triangle L} = \frac{\beta}{\alpha} \tag{12}$$

donde $\triangle$ son incrementos finitos en las variables W y L. La ecuación (12) impone una relación entre los incrementos de las posibles pérdidas y ganancias y las creencias (α, β) tal que mantiene constante el valor esperado de los pagos monetarios. Notar que si aumenta $\triangle L$ necesariamente ha de aumentar α para mantener la igualdad. Traducido a un lenguaje económico significa que si aumentan las posibles pérdidas han de aumentar las creencias sobre la probabilidad de éxito para mantener constante el valor esperado de los pagos monetarios. Nosotros nos centramos en el caso en que las variaciones se producen únicamente en W y L mientras las creencias iniciales se mantienen constantes.

En cuanto a la varianza, tenemos que (ver demostraciones en el Apéndice B.2):

$$Var(r \mid u) = \Big((W + \triangle W) - (L + \triangle L)\Big)^2 Var(y \mid u) \tag{13}$$

[3]Notar que para ambas figuras el resto de parámetros es idéntico.

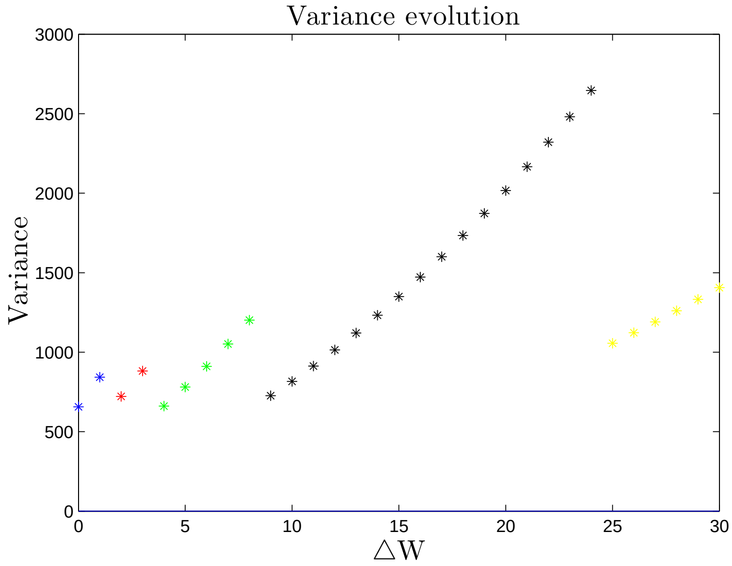

Figura 4: Incrementos en la varianza para $\alpha = 1$ y $\beta = 1$ fijos. Los valores $(\triangle W, \triangle L)$ mantienen constante el valor esperado de los pagos monetarios $r(t)$. Valores iniciales: $x(0) = 40$, $W = 10 + \triangle W$, $L = 5 + \triangle L$, $\rho = 0{,}01$ y $\lambda = 0{,}9$.

donde

$$Var(y \mid u) = u\left(\frac{\alpha}{\alpha+\beta} - \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}\right) + u^2\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{14}$$

La ecuación (13) es la varianza asociado a los incrementos $\triangle W$ y $\triangle L$ en W y L, respectivamente. Esto es, dados unos incrementos que preservan el valor esperado de la inversión, la varianza asociada a cada uno de los valores $W = 10 + \triangle W$ y $L = 5 + \triangle L$. Como podemos observar en la Figura 4, los incrementos tienen una consecuencia en la elección de la política óptima. Las discontinuidades que se aprecian son debidas a cambios en el numero de compras $u(0)$ óptimo, que en este ejemplo son (por colores y de izquierda a derecha): $u = 5, 4, 3, 2, 1$.

Si $\alpha = 1$ y $\beta = 1$ están fijados, siempre tenemos que $\mid \triangle W| = | \triangle L|$, lo que supone que tanto las pérdidas como las ganancias aumentan de igual forma respecto al centro de la distribución de los rendimientos esperados, manteniendo la media constante y aumentando la varianza. En ella se observa un crecimiento potencial de la varianza de acuerdo con el primer término de la derecha de la ecuación (13). Es importante notar que el paso en la política óptima de comenzar comprando 2 activos a 1, no supone un descenso en la varianza de la mitad, sino algo mayor de la mitad. Esto se debe a que la ecuación (14) no es lineal en u, lo que significa que hay un término cruzado cuando $u \geq 1$ que aumenta la varianza. Intuitivamente significa que cuando entran en juego dos o más activos, tenemos que tener en cuenta, no sólo lo que le sucede a cada uno de ellos por separado, sino también a la interacción entre ambos.

Volviendo a la política óptima del inversor, en esta sección se observa que para una inversión determinada la elección del *trader* también tiene en cuenta la varianza de la inversión: pasa de comenzar comprando un determinado u a otro menor cuando los incrementos de W aumentan (ver Figura 4), es decir, readapta la política cuando aumenta el riesgo. Este hecho, aunque lógico, es muy curioso porque en ningún momento se ha introducido explícitamente la varianza de la inversión en el algoritmo, pero la está teniendo en cuenta.

3.2.3. Influencia de las creencias iniciales: $\alpha(0)$ y $\beta(0)$

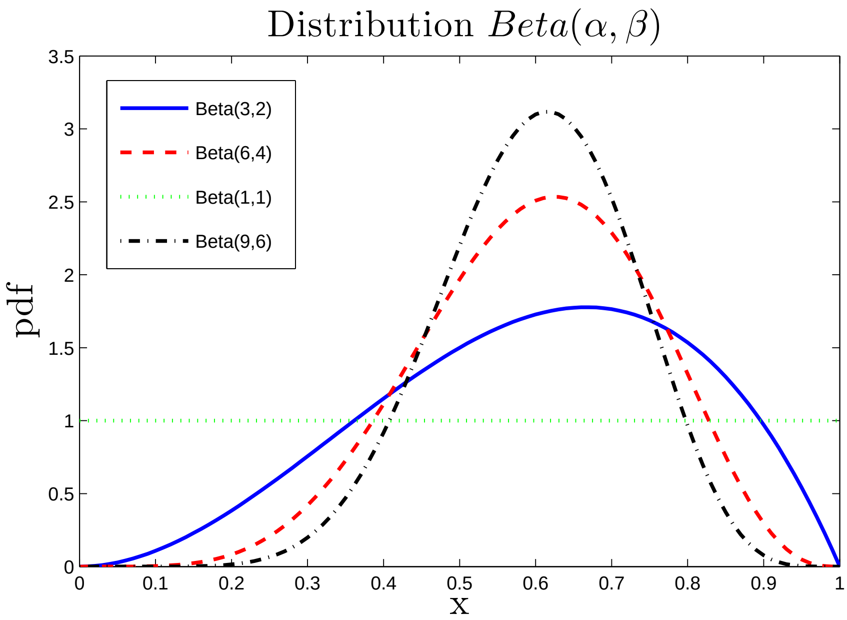

Figura 5: Bistribución *Beta* para distintos parámetros α, β

Si vamos a realizar cualquier inversión es por que tenemos algunas espectativas sobre cuál será la tendencia del mercado. Estas creencias inciales las podemos canalizar a traves de los parámetros (α, β). Como ya hemos dicho en repetidas ocasiones, modelizamos las creencias sobre la probabilidad de éxito mediante una distribucion $Beta(\alpha, \beta)$.

En los ejemplos anteriores hemos utilizado unos valores de las creencias iniciales relativamente desfavorable: aquellos en que $\alpha = \beta = 1$. Vemos en la Figura 5 que este caso corresponde a aquel en que no tenemos ninguna referencia acerca de cuál es la tendencia del mercado y comenzamos con una distribución uniforme en las creencias.

Las Figuras 6 y 7 representan las políticas óptimas cuando las creencias inciales son acertadas y cuando son erróneas. Para ello se escogen valores de (α, β) tales que el valor esperado de la *Beta* sea igual a δ, Figura 6, y cuando estamos lejos de él, Figura 7. Lo que observamos en el segundo caso es una política que de partida es mucho

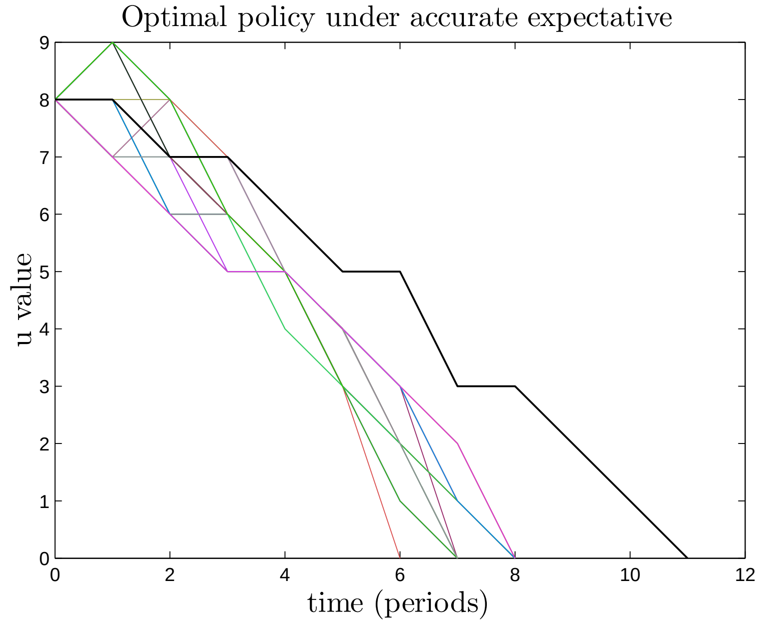

Figura 6: Política óptima bajo expectativas precisas. Valores inciales: $x(0) = 40$, $\boldsymbol{\alpha(0) = 7}$, $\boldsymbol{\beta(0) = 3}$, $W = 10$, $L = 5$, $\rho = 0{,}01$, $\lambda = 0{,}9$, $\delta = 0{,}7$. Número de iteraciones de Monte-Carlo: 100.

más conservadora, efecto debido a unas falsas creencias inciales poco favorables, y que conforme observa las reacciones del marcado va siendo menos conservadora: efecto de remontada en el número de compras debido al aprendizaje. Si nos fijamos en la Figura 1 podemos observar un comportamiento similar a este aunque menos pronunciado: en la Figura 1 las creencias inciales eran $(\alpha, \beta) = (1, 1)$ y en la Figura 7 $(\alpha, \beta) = (3, 7)$. Esto es una demostración de la capacidad de aprendizaje del algoritmo y, desde el punto de vista de las finanzas, permite explotar situaciones en las que a priori nunca lo hubiesemos hecho, precisamente por nuestras creencias erroneas. Este es un punto fuerte de este sistema con respecto al sistema de inversiones en un solo periodo: la secuenciación de las compras permite corregir las creencias iniciales maximizando los beneficios y la utilidad de todo el proceso. En el caso de la Figura 6 esto no sucede: como las creencias iniciales son acertadas, no hay una fase de "experimentación" en la que se aprende, sino que directamente se explota el conocimiento que se tiene.
Otro caso plausible es aquel en que nuestras creencias son favorables pero la realidad del mercado es la contraria. Ello se ilustra en la Figura 8. Lo que observamos es que el número de compras se retrae con respecto al comienzo debido al aprendizaje. El algoritmo aprende que las creencias iniciales no se corresponden con la realidad del mercado y modera las compras rápidamente.

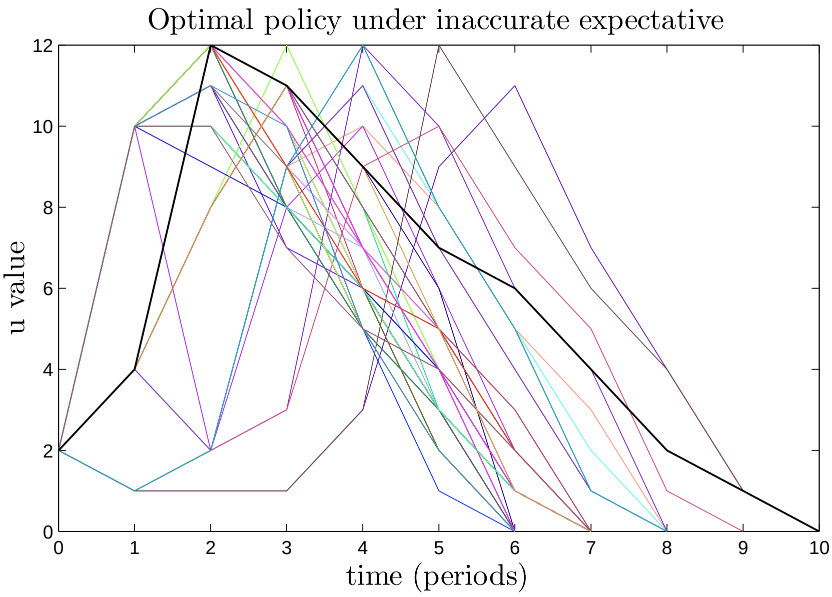

Figura 7: Política óptima bajo expectativas inprecisas. Valores inciales: $x(0) = 40$, $\boldsymbol{\alpha(0) = 3}$, $\boldsymbol{\beta(0) = 7}$, $W = 10$, $L = 5$, $\rho = 0{,}007$, $\lambda = 0{,}9$, $\delta = 0{,}7$. Número de iteraciones de Monte-Carlo: 100.

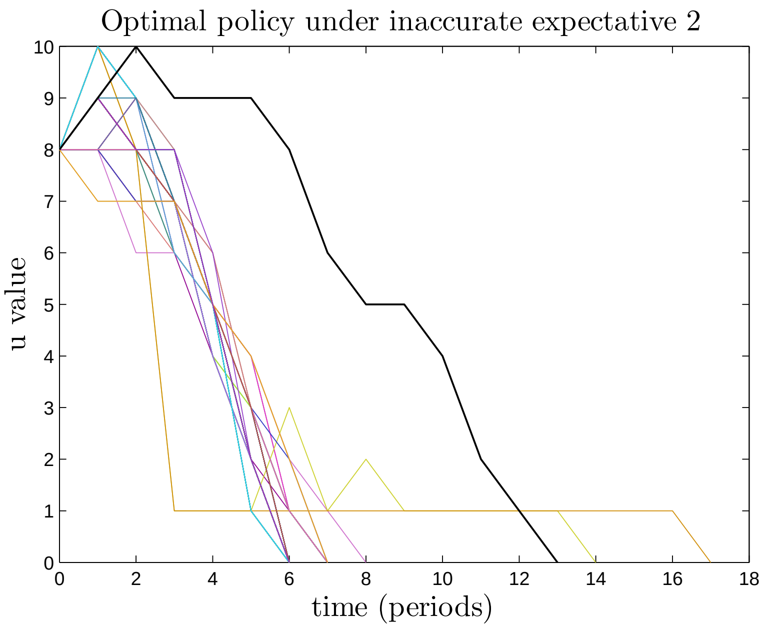

Figura 8: Política óptima bajo expectativas inprecisas 2. Valores inciales: $x(0) = 40$, $\boldsymbol{\alpha(0) = 7}$, $\boldsymbol{\beta(0) = 3}$, $W = 10$, $L = 5$, $\rho = 0{,}01$, $\lambda = 0{,}9$, $\boldsymbol{\delta = 0{,}45}$. Número de iteraciones de Monte-Carlo: 100.

4. Conclusiones

Existe un gran número de situaciones en las que la interacción entre las decisiones tomadas y las posibilidades de aprendizaje devenidas deben ser sopesadas. Este trabajo presenta una solución basada en algoritmos de optimización dinámica y la estadísitca Bayesiana capaz de resolver de manera exacta el *trade-off* entre la maximización de la utilidad esperada y la capacidad de aprendizaje en el contexto de horizontes temporales finitos. Mediante estas técnicas se presenta una herramienta útil para la determinación de políticas óptimas en la compra de activos financieros, secuenciando las compras en el tiempo para aprender en el proceso y maximizar la utilidad esperada y los beneficios.

El método es tranferible a muy diversas aplicaciones, tanto en el contexto de las finanzas, como de los seguros o las políticas óptimas de nuncios. Es de especial interés la generalización del problema a distintos tipos de activos. Los algoritmos aquí mostrados son extensibles en esta línea. Otro aspecto interesante es analizar las posibilidades de las políticas subóptimas: éstas no son exactas pero tienen muchas ventajas computacionales en cuanto a la velocidad de ejecución.

Apéndice

A. Algunos detalles técnicos

A.1. Sobre la función de utilidad

Para el cálculo de la función de utilidad (3) que se muestra en la sección 2.2, el algoritmo contiene la aproximación a segundo orden del desarollo de Taylor centrado en cero.

La reescribimos (3) como:

$$E(-e^{\rho r}) \simeq -1 + \rho E(r) - \frac{1}{2}\rho^2 E(r^2) \tag{15}$$

donde:

$$E(r \mid u) = E\big(Wy + L(u-y)\big) = E\big((W-L)y + Lu\big) = (W-L)E(y \mid u) + Lu \tag{16}$$

$$\begin{aligned} E(r^2 \mid u) &= E\left\{\Big((W-L)y + Lu\Big)^2\right\} = ... \\ &= (W-L)^2 E(y^2 \mid u) + L^2u^2 + 2L(W-L)uE(y \mid u) \end{aligned} \tag{17}$$

A.2. Sobre distribuciones, momentos y probabilidad

A continuación se muestra la distrubucíon de las variables así como sus *momentos* de primer y segundo orden.

$\boldsymbol{y \sim Binomial(\delta, u)}$	$\boldsymbol{\delta \sim Beta(\alpha, \beta)}$
$\mathrm{E}\{\mathrm{y} \mid \delta\} = u\delta$	$E\{\delta\} = \frac{\alpha}{\alpha+\beta}$
$E\{y^2 \mid \delta\} = u\delta + u(u-1)\delta^2$	$E\{\delta^2\} = \frac{\alpha(\alpha+\beta)}{(\alpha+\beta)(\alpha+\beta+1)}$
$Var\{y \mid \delta\} = u\delta(1-\delta)$	$Var\{\delta\} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Finalmente, la probabilidad en (10) se calcula a partir de la distribución *Beta*, tal que:

$$\Pr(\hat{\gamma} \mid x, u, \gamma) = \Pr(y \mid \alpha, \beta, u) = \int_0^1 \Pr(y \mid \delta, u)\phi(\delta; \alpha, \beta) d\delta$$

donde $\phi(\cdot; \alpha, \beta)$ es la densidad de la función de distribución de probabilidad $Beta(\alpha, \beta)$. Y $\Pr(y \mid \alpha, \beta, u)$ es:

$$\Pr(y \mid \alpha, \beta, u) = \binom{u}{y} \frac{B(\alpha + y, \beta + u - y)}{B(\alpha, \beta)}$$

donde $B(\Delta)$ denota la función Beta.

B. Mean-preserving Spread

B.1. Valor Esperado constante

Dada la ec. (2), supongamos que $r(t)$ es constante,

$$dr = dWy + (u - y)dL = 0$$

donde sin pérdida de generalidad se han omitido las dependencias temporales. Reorganizando y tomando esperanzas condicionales:

$$(dW - dL)E(y \mid u) + udL = 0$$

expresión que podemos escribir como:

$$1 - \frac{dW}{dL} = \frac{u}{E(y \mid u)}$$

La *Ley de Esperanzas Iteradas* aplicada sobre y:

$$E(y \mid u) = E\Big(E(y \mid \delta) \mid u\Big) = E\left(u\delta \mid u\right) = uE\left(\delta\right) = u\frac{\alpha}{\alpha + \beta}$$

donde el valor esperado de las distribuciones están especificados en el Apéndice A.2.

Introduciendo la última ecuación en la penúltima llegamos a la ecuación diferencial:

$$-\frac{dW}{dL} = \frac{\beta}{\alpha}$$

e integrándola entre $\int_{W_i}^{W_f}$ y $\int_{L_i}^{L_f}$, respectivamente[4], obtenemos las relación entre W, L, α y β que mantienen el valor esperado constante:

$$\boxed{-\frac{\triangle W}{\triangle L} = \frac{\beta}{\alpha}}$$

que es la ec. (12) que queríamos demostrar.

[4]Como las pérdidas L han sido definidas como positivas hay un cambio de signo en su integral

B.2. Varianza

En esta ocasión buscamos la influencia sobre la varianza de los incrementos $\triangle W$ y $\triangle L$. De esta manera buscamos la varianza de:

$$r(t) + \triangle r(t) = y(t)\Big(W + \triangle W\Big) + \Big(u(t) - y(t)\Big)\Big(L + \triangle L\Big)$$

como $\triangle W$ y $\triangle L$ son tales que $\triangle r(t) = 0$ Reorganizando y tomando la varianza:

$$Var(r \mid u) = \Big((W + \triangle W) - (L + \triangle L)\Big)^2 Var(y \mid u)$$

Aplicando la *Ley de la Varianza Total* sobre y:

$$\begin{aligned} Var(y \mid u) &= E\Big(Var(y \mid u)\Big) + Var\Big(E(y \mid u)\Big) \\ &= E\Big(u\delta(1-\delta)\Big) + Var\Big(u\delta\Big) \\ &= uE(\delta) - uE(\delta^2) + u^2 Var(\delta) \\ &= u\Big(E(\delta) - E(\delta^2)\Big) + u^2 Var(\delta) \end{aligned}$$

explícitamente en función de α y β según Apendice A.2:

$$Var(y \mid u) = u\left(\frac{\alpha}{\alpha+\beta} - \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}\right) + u^2 \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Referencias

[1] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*, volume I. Athena Scientific, 2005.

[2] Francisco Alvarez; Alfredo Garcia-Hiernaux. Bounded expreimentation in a calibrated model of adversing. Working Paper, October 2012.

[3] Siu Lung Law. *Financial Optimization Problems.* PhD thesis, St. Anne's Collage. University of Oxford, 2005.

[4] Dimitris Bertsimas; Dessislava Pachamanova;. Robust multiperiod management in the presence of transaction costs. *Computers and operations research*, (35):3–17, 2008.

[5] Nikolas Topaloglou; Hercules Vladimirou; Stavros A. Zenios;. A dynamic stochastic programming model for international porfolio management. *European journal of operational research*, 2006.