

VIABILIDAD DE UN PLL (PHASE-LOCKED LOOP) PARA LA EXTRACCIÓN DE COMPONENTES SINUSOIDALES EN SERIES DE DATOS FINANCIEROS.

Antonio Mínguez Olivares

Departamento de Ingeniería Audiovisual y Comunicaciones.
EUIT Telecomunicación, Universidad Politécnica de Madrid. España.
aminguez@diac.upm.es, +34-913367794

RESUMEN: Las series de datos financieros (SP500, Dow Jones, Nasdaq, etc) se pueden analizar aplicando técnicas de procesamiento digital de señales con objeto de extraer cierta información. En los últimos años ha aparecido una nueva técnica que permite descomponer una señal en sus cíclicas, la transformada de Hilbert-Huang. Cuando se aplica esta transformada a una serie de datos financieros aparecen separadas varias componentes sinusoidales y una componente tendencial. Tomando como referencia cualquiera de las componentes sinusoidales y aplicando simples señales de trading sobre la misma es factible obtener rentabilidades positivas de forma continuada. Sin embargo, estas señales están generadas a posteriori. Con objeto de obtener estas señales en tiempo real (online) se propone la utilización de un PLL (Phase-Locked Loop) para tener una referencia enganchada en fase en cualquiera de las componentes sinusoidales, obtenidas previamente con la transformada de Hilbert-Huang, y a partir de ella realizar una estrategia de inversión.

PALABRAS CLAVE:

Descomposición Modal Empírica EMD, *Phase-Locked-Loop* (PLL), Transformada de Hilbert-Huang.

1 INTRODUCCIÓN

Las técnicas de procesamiento digital de señales [1], utilizadas habitualmente en disciplinas afines a la Ingeniería de Telecomunicación, nos permiten analizar las fluctuaciones de las cotizaciones de precios que se producen en los mercados financieros desde un punto de vista diferente. Estas fluctuaciones de precios de compra-venta quedan registradas en series temporales de datos de una forma similar a las series de datos procedentes del muestreo de señales analógicas, de acuerdo con el teorema de *Nyquist* [2]. Así, una serie de datos de un valor cotizado en bolsa, por ejemplo de Microsoft (que cotiza con el *ticker* MSFT), nos muestra los precios de su cotización a intervalos regulares de tiempo: desde intervalos regulares de minutos hasta intervalos incluso de meses. Este intervalo corresponde al periodo de muestreo, según la terminología utilizada en el propio teorema de *Nyquist*.

En este estudio vamos a utilizar series temporales con datos diarios, es decir, un dato de cotización por día y justamente el dato de cierre de la sesión. Además, en

vez de utilizar un valor concreto como Microsoft, utilizaremos un índice tan significativo como el *Standard & Poor's 500 Index* (índice Standard & Poor's 500), también conocido como S&P 500. Es uno de los índices bursátiles más importantes de Estados Unidos y es considerado como el índice más representativo de la situación real del mercado. Sus datos históricos de pueden obtener directamente desde página financiera de Yahoo [3]. En la figura siguiente se puede apreciar la variación de los precios de este índice desde el año 2000.

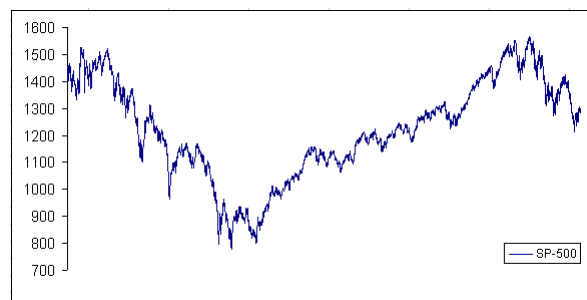


Figura1. Representación gráfica de los precios de cierre del índice SP500. Periodo 2000-2008.

Es obvio que, para obtener una rentabilidad positiva hay que comprar "barato" y vender "caro", es decir, comprar cuando un valor está a bajo precio y vender cuando el precio está alto, aunque no necesariamente en este orden. Hoy en día existen productos financieros, como los "futuros", que permiten vender antes de comprar. La justificación y explicación de estos productos financieros [4] se sale fuera del alcance de este artículo. En definitiva, se pueden obtener rentabilidades positivas tanto cuando la bolsa sube como cuando la bolsa baja, claro está si uno sabe de antemano como se va a comportar la misma, o dicho de una forma más formal, si es capaz de predecir su comportamiento.

Aquí es donde entran en juego las técnicas de procesamiento de señales. Estas nos van a permitir analizar el tipo de datos con el que nos estamos enfrentando. Si se tratara de una señal aleatoria pura cualquier estrategia de compra-venta sería puro azar y no se diferenciaría mucho de la que por ejemplo pudiéramos aplicar en un casino de Las Vegas. Pero si dentro de la serie de datos existieran componentes periódicas o cíclicas, entonces para obtener una rentabilidad positiva

sería tan sencillo como comprar en la parte baja del ciclo (cuando el precio está bajo) y vender en la parte alta del ciclo (cuando el precio está alto).

¿Cómo saber entonces si existen componentes cíclicas en una serie de datos? En una primera aproximación bastaría con hacer un simple análisis espectral con la transformada de Fourier [1] que nos presentara dicha información en el dominio de la frecuencia. Sin embargo, el análisis utilizando esta transformada nos llevaría a un compromiso de resolución tiempo-frecuencia, en la que nos sería imposible determinar con precisión el valor de las componentes cíclicas en cada momento para poder aplicar, en tiempo real, cualquier estrategia de *trading*.

Para hacer este análisis es mucho más efectivo utilizar una novedosa técnica de descomposición conocida por la "Transformada de Hilbert-Huang" [5] en la que se incluye la denominada "descomposición modal empírica" (*Empirical Mode Decomposition* -EMD-) [6]. Utilizando esta técnica es posible extraer cada una de las componentes cíclicas que existen en la serie de datos, además de la componente tendencial (componente continua o D.C.), de una forma óptima, siguiendo las variaciones instantáneas de frecuencia que presenten cada una de las componentes cíclicas. Esto es de suma importancia puesto que permitirá saber en todo momento en qué fase del ciclo nos encontramos para aplicar de forma correcta una estrategia de compra-venta. Desafortunadamente, esto no será factible en la práctica debido a ciertas limitaciones de la descomposición EMD [6], pero la información de qué componentes cíclicas han existido en la serie histórica de datos nos permitirá focalizar el problema abordándolo desde el punto de vista de "extracción de componentes sinusoidales" de una señal. Algunas alternativas pueden ser:

- Filtrado paso banda (FIR o IIR) [1]
- Filtrado adaptativo [7]
- PLL (*phase locked loop*) enganchado en fase con la componente sinusoidal deseada. [8]

El resto del artículo está organizado de la siguiente manera. Primero se explicará brevemente la descomposición modal empírica y su aplicación a la serie histórica de datos diarios del SP500. Se estudiarán las rentabilidades máximas que se pueden alcanzar en las componentes cíclicas presentes. Posteriormente se estudiará el diseño de un algoritmo PLL para extraer una componente cíclica cualquiera y analizar cómo se altera la rentabilidad basada en esa componente cíclica. El artículo resumirá en el apartado de conclusiones la viabilidad o no de utilizar un PLL para extraer una componente sinusoidal en base a los resultados obtenidos.

2 DESCOMPOSICIÓN MODAL EMPÍRICA

Las señales procedentes de fenómenos naturales son de origen no-estacionario y no-lineal. En 1998 N.E. Huang [6] desarrolló una nueva técnica para representar dichas señales de forma adaptativa, como la suma de componentes cíclicas moduladas en amplitud y frecuencia, mas un residuo o componente tendencial. A este proceso lo denominó *Empirical Mode Decomposition* (EMD) y a cada una de las componentes *Intrinsic Mode Functions* (IMFs). La transformada Hilbert aplicada a cada una de las IMFs y la posterior obtención del valor de la frecuencia instantánea se conoce con el nombre de Transformada de Hilbert-Huang (HHT). La transformada HHT permite realizar un análisis espectral de señales de origen no-estacionario y no-lineal utilizando un algoritmo que procesa los datos en su conjunto (a posteriori), el algoritmo EMD. El análisis espectral obtenido con la transformada HHT tiene una resolución óptima, superior a cualquiera de las técnicas tradicionales y puede aplicarse, como se ha mencionado anteriormente, a señales no-estacionarias, con variaciones dinámicas de amplitud y frecuencia.

Básicamente, la descomposición EMD parte de la propia forma de onda de los datos para ir extrayendo directamente en el dominio del tiempo, de una forma iterativa, cada una de las IMFs de las que está compuesta la señal. Identifica los puntos máximos y mínimos ("picos"), y considera que debe existir una componente IMF (una oscilación) montada sobre el valor de la media local de la señal. Las condiciones que deben aplicarse para extraer cada componente oscilatoria son:

1. El número de "picos" y el número de cruces con la media local han de ser iguales o diferir como máximo en 1.
2. En cualquier punto, el valor medio de la envolvente que une los "máximos" y de la envolvente que une los "mínimos" ha de ser nulo.

Por tanto, el algoritmo EMD lo que hace es expresar una señal $x(t)$ como la suma de una oscilación más una componente de baja frecuencia o residuo. Una vez separadas estas dos componentes, se vuelve a aplicar el proceso al residuo considerándolo como la nueva señal, y así sucesivamente hasta que se incumple la primera de las dos condiciones expuestas anteriormente, en cuyo caso el residuo resultante corresponderá a parte tendencial de la señal. En definitiva, el algoritmo se puede resumir de la siguiente forma:

1. Identificación de los máximos y los mínimos de la forma de onda $x(t)$
2. Envolvente de los máximos mediante interpolación $e_{\min}(t)$
3. Envolvente de los mínimos mediante interpolación $e_{\max}(t)$
4. Cálculo de la media $m(t) = [e_{\max}(t) + e_{\min}(t)] / 2$
5. Obtención de la IMF $k_i = x(t) - m(t)$

6. Repetición el proceso sobre k_i , es decir, $k_i = x(t)$ y volver al punto 1.

Los puntos 1-4 se repiten en cada iteración hasta que se alcanza una media de valor nulo, tal como se establece en la segunda condición. A este proceso se le denomina *sifting*. La interpolación que se suele aplicar aquí es de orden tres, más que suficiente para obtener resultados coherentes, aunque es un tema abierto de estudio.

3 APLICACIÓN DEL ALGORITMO EMD A LA SERIE DE DATOS DIARIA DEL SP500

La aplicación del algoritmo EMD a la serie de datos diarios del índice SP500 nos permite apreciar cada una de las componentes cíclicas de las que está formada. Los datos diarios de cierre de este índice, como ya se dijo en un principio, se han obtenido de la página financiera de Yahoo [3]. La serie histórica comprende desde el 01-03-2000 hasta el 18-08-2008. Como puede intuirse en la figura 1, la serie empieza con el valor 1455.22 y termina con el valor 1278.6, habiendo alcanzado un máximo de 1565.15 el día 09-10-2007 y un mínimo de 776.76 el día 09-10-2002. El número de componentes IMFs obtenidas en la descomposición han sido 14. La IMF-14 corresponde al residuo final o componente tendencial. Cuanto más pequeño es el número de la IMF más alta es la frecuencia que representa, por tanto la componente IMF-1 podemos considerar que es una componente casi aleatoria, de muy alta frecuencia y el resto, desde la IMF-2 hasta la IMF-14, son las distintas componentes cíclicas que forman parte de la serie de datos.

A continuación se presentan las gráficas de estas componentes. En primer lugar las IMF más altas, desde la 6 a la 14, para toda la serie histórica de datos. El resto de componentes, las de más alta frecuencia, se mostrarán en la figura posterior sólo para un intervalo de la serie histórica -2007-, con objeto de que se puedan apreciar visualmente el carácter sinusoidal, localizado en ese intervalo (*zoom*).

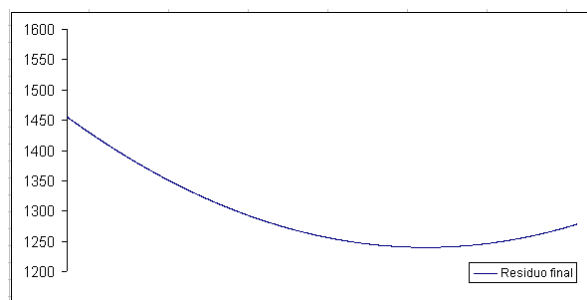
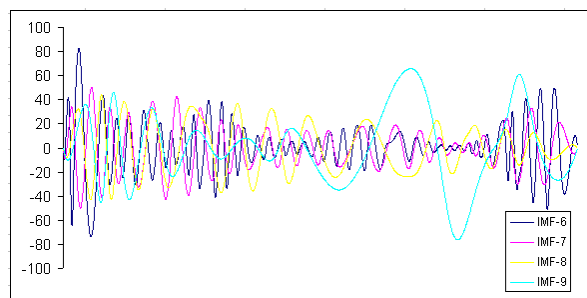
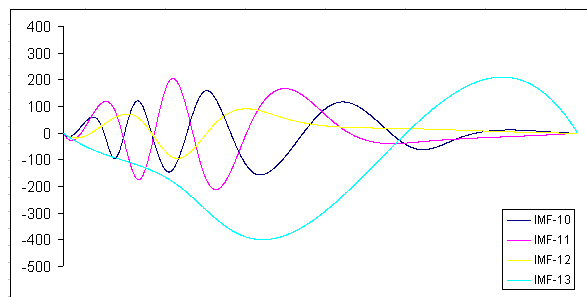
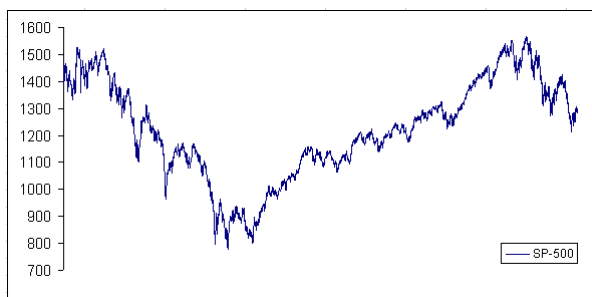
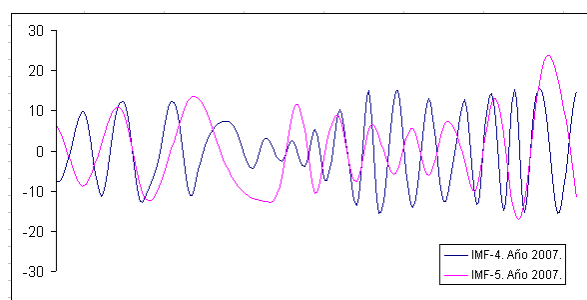
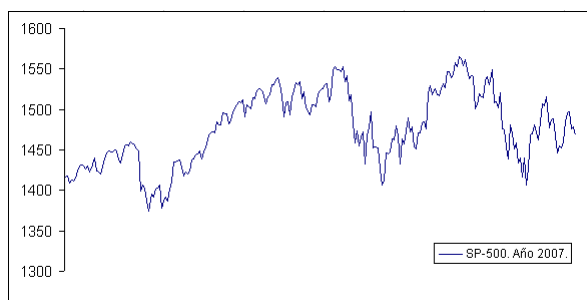


Figura 2. Serie histórica diaria del SP500 e IMFs (6-14). Período 2000-2008.



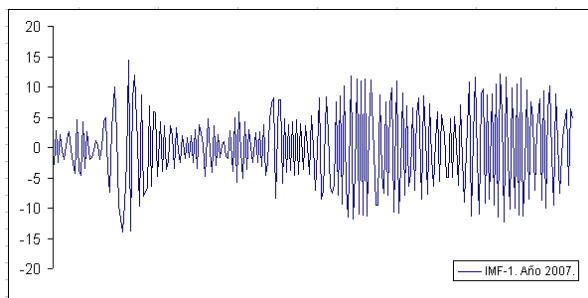
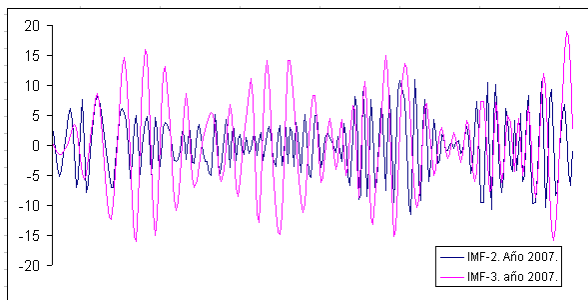


Figura 3. Serie histórica diaria del SP500 e IMFs (1-5). Periodo 2007.

Si sobre estas componentes cíclicas se construye una simple estrategia de compra-venta, de modo que cuando el dato de una IMF sea mayor que el anterior se produzca una orden de compra (tramo alcista del IMF) y cuando el dato de la IMF sea inferior que el anterior se produzca una orden de venta (tramo bajista de la IMF), las rentabilidades que se obtienen son excepcionalmente buenas, tal como se aprecia en la tabla 1. En ella se muestran las rentabilidades anuales obtenidas utilizando como señales de compra y venta los cambios de fase en cada una de las componentes cíclicas obtenidas en la descomposición EMD.

Las mayores rentabilidades se obtienen en las IMFs más bajas, componentes de más alta frecuencia y que producen un mayor número de órdenes de compra y venta. No se muestran las rentabilidades de las últimas 4 IMFs (10-13) por ser prácticamente nulas o negativas al haber poquísimas señales de compra-venta. En el extremo opuesto, la componente IMF-1 puede ser considerada como una componente casi aleatoria y la estrategia de compra-venta no funciona de igual forma que en el resto de componentes, aunque también es posible obtener muy buenos resultados de rentabilidad sobre ella con otro tipo de estrategia.

Año	IMFs. Rentabilidades porcentuales (%).								
	2	3	4	5	6	7	8	9	
2000	40	87	45	17	34	-15	16	-4	
2001	66	96	51	76	43	55	27	7	
2002	50	90	96	71	40	9	21	23	
2003	45	56	59	17	24	24	27	-12	
2004	30	23	34	36	33	9	14	-3	
2005	35	35	31	14	25	26	8	10	
2006	26	49	27	17	10	7	8	-6	
2007	41	63	48	9	41	14	-2	11	
2008	34	32	29	25	-6	18	2	3	
Total	371	533	424	286	247	149	124	30	

Tabla 1. Rentabilidades anuales porcentuales obtenidas a partir de las componentes IMFs de la descomposición EMD sobre la serie histórica de datos diarios del índice SP500.

Hay que resaltar que estos datos no pueden ser considerados como reales ya que las IMFs se han obtenido a posteriori, es decir, con toda la serie de datos históricos. Para que estas rentabilidades fueran ciertas cada componente IMF debería ser calculada de forma "on-line". Sin embargo, la descomposición EMD nos permite tener una estimación de las frecuencias presentes en la serie histórica de datos para intentar extraer cualquiera de ellas mediante un proceso alternativo. Algunos de ellos se indicaron anteriormente:

1. Por filtrado paso-banda centrado en el margen de frecuencias en los que se encuentra la correspondiente IMF.
2. Por filtrado adaptativo, que siga las variaciones de frecuencia de una determinada IMF
3. Mediante un PLL que se enganche en fase a las variaciones dinámicas de la IMF

En este artículo se aborda este último procedimiento con objeto construir una componente sinusoidal y que nos permita obtener a través de ella las correspondientes órdenes de compra-venta y su posterior análisis de rentabilidad.

4 SISTEMA PLL (PHASE-LOCKED-LOOP)

En principio la implementación de un PLL puede ser utilizado para generar cualquiera de las sinusoides presentes en la serie de datos. En este caso concreto, se utilizará para regenerar la componente IMF de más alta frecuencia, es decir, la IMF-1. El análisis detallado de las variaciones de frecuencia de dicha componente nos indica que su margen de variación (teniendo en cuenta que la frecuencia de muestreo normalizada es 1) oscila entre 1/2 y 1/6. Por tanto, el algoritmo PLL deberá engancharse dentro de este margen.

Un sistema PLL (*phase-locked-loop*) es un sistema de lazo cerrado que permite que una señal de referencia externa controle la frecuencia y la fase de un

oscilador. Los PLL se usan básicamente en sistemas de telecomunicación para: generar portadoras de modulación, generación de osciladores locales en recepción, sintetizadores de frecuencia, demoduladores, recuperación de impulsos de reloj en transmisiones digitales, etc.

El diagrama de bloques de un sistema PLL se muestra en la figura 4. En el se pueden apreciar cada uno de sus elementos. La referencia externa y la señal del oscilador local se comparan en un detector de fase. La salida del detector de fase pasa por un filtro paso bajo y la salida de este es una señal de control que gobierna la frecuencia del oscilador controlado por tensión (VCO).

El detector de fase se comporta como un mezclador de la frecuencia externa (f_s) y de la frecuencia del VCO (f_o), entregando en su salida la mezcla de ambas $[f_s - f_o]$. Los productos de alta frecuencia $f_o + f_s$, $2f_s$, $2f_o$, etc. son eliminados por el filtro paso bajo. La tensión de salida del filtro paso bajo, de frecuencia $[f_s - f_o]$, controlará el VCO hasta que ambas frecuencias sean iguales. Cuando esto ocurre, la tensión de salida del filtro paso bajo se estabiliza con una componente continua y la diferencia de fases entre la señal externa y la señal del VCO se mantiene constante.

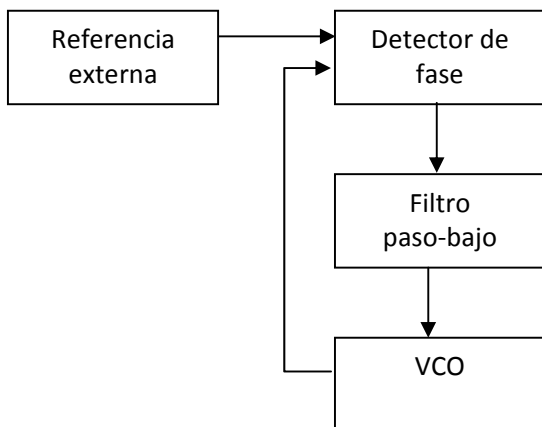


Figura 4. Diagrama de bloques de un sistema PLL (phase-locked-loop)

En el caso que nos ocupa hay que señalar que las señales que manejamos son realmente secuencias numéricas y por tanto los elementos del diagrama de bloques son algo diferentes. El detector de fase es digital, el filtro paso bajo es un sistema discreto FIR y el VCO es reemplazado por un NCO (oscilador controlado numéricamente). Todo el sistema ha sido implementado en el entorno de programación de MATLAB [9].

La señal de referencia, señal histórica de datos se acondiciona, antes de ser aplica al PLL digital, con un simple filtro paso alto digital de la forma: $y(n) = [x(n) - x(n-1)]/2$, es decir, los datos de referencia son la semidiferencia de los dos últimos datos de la serie histórica en cada momento. De esta forma atenúamos el resto de componentes IMFs que no van a ser objeto de

análisis. En la figura 5 se presenta la forma de onda de la secuencia de datos de referencia así calculada, concretamente 251 datos del año 2007 (datos filtrados paso-alto de la secuencia histórica del SP500 correspondientes a 251 días de negociación).

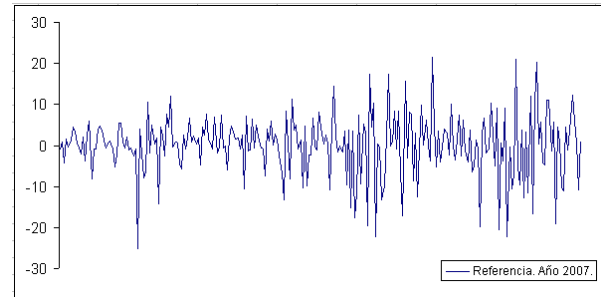


Figura 5. Señal de referencia. Entrada al PLL digital.

En la figura siguiente se muestra la señal generada por el oscilador digital del sistema PLL. Es una señal alterna de dos estados, suficiente para producir las correspondientes señales de compra-venta.

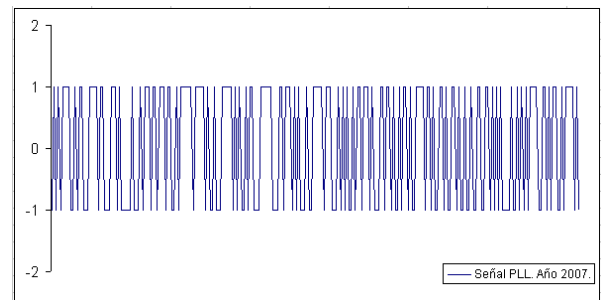


Figura 6. Señal generada por el oscilador del sistema PLL.

Una vez obtenida la señal del oscilador del PLL lo que hay que comprobar es la bondad de las órdenes de compra-venta generadas a partir de la misma. Esta bondad se mide en función de la rentabilidad anual porcentual generada y comparada con la ideal supuestamente generada con la componente IMF-1, que es a la que debe haberse enganchado en fase el sistema PLL. En la tabla siguiente se muestra una comparación de las rentabilidades obtenidas con cada una de estas señales.

Año	Rentabilidades (%)	
	IMF-1	PLL-Digital
2000	85.96	0.21
2001	75.24	4.52
2002	136.03	17.11
2003	105.11	33.08
2004	45.34	3.87
2005	39.73	19.83
2006	50.36	2.22
2007	102.44	41.02
2008	79.97	41.85
Total	719.28	163.69

Tabla 2. Tabla de rentabilidades anuales porcentuales.

Como puede observarse, las rentabilidades obtenidas con la componente IMF-1 son espectaculares pero no son reales puesto que se han obtenido con datos "a posteriori". Las obtenidas con la señal del oscilador del PLL obviamente son sensiblemente inferiores pero son aceptables. En ningún año se obtienen rentabilidades negativas y en promedio, la rentabilidad anual es superior al 18%.

5 CONCLUSIONES

En este artículo se ha estudiado la viabilidad de utilizar un sistema PLL (*phase-locked-loop*) para extraer componentes cíclicas de una serie de datos financieros, concretamente la del índice SP500. La certeza de la existencia de estas componentes cíclicas se ha obtenido a través de la descomposición EMD utilizada en la transformada de Hilbert-Huang. El análisis de rentabilidad se ha centrado en la primera componente de la descomposición EMD, la de mayor frecuencia, y los resultados obtenidos demuestran que es posible acumular rentabilidades aceptables año tras año tomando como base únicamente la información de fases de la componente periódica del oscilador del PLL. Aunque las rentabilidades no son uniformes, en cinco de los nueve años estudiados superan el 17%.

6 REFERENCIAS

- [1] Oppenheim, A. V. and R. W. Schaffer, Discrete-Time Signal Processing, Prentice-Hall, 1989.
- [2] H. Nyquist, "Certain topics in telegraph transmission theory," Trans. AIEE, volumen 47, páginas 617-644, abril 1928.
- [3] Datos históricos de SP500.
(<http://es.finance.yahoo.com/q?s=^GSPC>)
- [4] Hull, John C. (2005), Options, Futures and Other Derivatives (excerpt by Fan Zhang) (6th ed.), Prentice-Hall, ISBN 0131499084.
- [5] N. E. Huang, Z. Shen, Hilbert-Huang Transform and Its Applications, Ed. World Scientific Co. Pte. Ltd. 2005. ISBN 981-256-376-8.
- [6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, y H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. Lond. A, vol. 454, pp. 903-995, 1998.
- [7] B. Widrow and S. D. Stearns. Adaptive Signal Processing. New Jersey: Prentice-Hall, Inc., 1985.
- [8] William F. Egan, Frequency Synthesis by Phase Lock (2nd ed.), John Wiley and Sons, 2000
- [9] <http://www.mathworks.com/products/matlab/>