

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE
INGENIEROS DE TELECOMUNICACIÓN



PROYECTO FIN DE CARRERA

DISEÑO E IMPLEMENTACIÓN DE UN SIMULADOR PARA SISTEMAS ALGORÍTMICOS DE TRADING: PREPROCESADO DE DATOS, CÁLCULO DE MEDIDAS DE RENDIMIENTO Y SOBREOPTIMIZACIÓN

Autor: D. Antonio J. Turel Rodríguez

Tutor: Dr. Eduardo López Gonzalo

Departamento: Señales, Sistemas y Radiocomunicaciones

Miembros del tribunal:

Presidente: Dr. Jesús Alcázar Fernández

Vocal: Dr. Luis Alfonso Hernández Gómez

Secretario: Dr. Eduardo López Gonzalo

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE
INGENIEROS DE TELECOMUNICACIÓN

Licenciatura en Ingeniería de Telecomunicación



PROYECTO FIN DE CARRERA

**DISEÑO E IMPLEMENTACIÓN DE UN SIMULADOR
PARA SISTEMAS ALGORÍTMICOS DE TRADING:
PREPROCESADO DE DATOS, CÁLCULO DE
MEDIDAS DE RENDIMIENTO Y
SOBREOPTIMIZACIÓN**

D. ANTONIO J. TUREL RODRÍGUEZ

2014

Resumen

Actualmente el sector financiero sigue adaptándose al trading algorítmico, cada vez más tecnológico como es el *High Frequency Trading* (HFT). Es decir, las empresas tecnológicas que aportan nuevas bases de datos con lecturas/escrituras más rápidas y nuevos algoritmos para tratar BIG DATA, entre otros, son muy necesarias en los Mercados Financieros. La operativa discrecional decrece su interés mientras aumentan los contenidos en computer science como es el análisis cuantitativo.

Durante el trabajo se nombra al equipo necesario en las empresas financieras de última generación, formado por varios roles y éstos principalmente técnicos. De forma muy resumida podríamos decir que los *traders* aportan la idea, los *quants* formalizan el modelo de la idea y los de tecnología mantienen la conexión con el mercado para que el modelo continúe su ejecución normal. Es decir, se está reclutando principalmente a personal cualificado en ciencias y tecnología (*Quant, IT*) en los despachos financieros por aquello de obtener resultados con fundamento científico, más que la figura de lo que era el *trader*.

Ya no hablamos sólo del hecho de poder trabajar en uno de los sectores en expansión sino que los mercados financieros se han automatizado por completo y ello implica la facilidad (PC + internet) del inversor particular, aquel que tiene la oportunidad de gestionar de manera eficiente sus ahorros, de poder conectarse al mercado con una cuenta contratada con un intermediario financiero.

El reto que se plantea una persona que se conecta a los mercados es la capacidad de análisis con su metodología e implementación de estrategias pero para ello hay que ser responsable del capital seleccionado para invertir entre todos los ahorros, en qué condiciones se puede invertir (características del derivado/intermediario financiero y posible *portfolio*), cuánta rentabilidad se estima obtener y cuánto riesgo podemos asumir de pérdida. Ser riguroso en una metodología sistemática es muy importante para cumplir con los rendimientos estimados y comprobar mediante la estadística/probabilidad si los resultados son coherentes o no, porque hacer un ‘traje a la medida’, como dice mi abuelo, de la estrategia analizada en el pasado lo normal será que aparezca el ‘coco’ de la sobreoptimización y con ello unas pérdidas cuando ejecutamos en tiempo real que no deseábamos obtener.

Para conseguir liderar los mercados financieros, lógicamente se necesita una cartera grande pero también una posible inversión en infraestructura computacional. Porque tenemos en este mundo varias ligas: inversión a largo, medio plazo, diario, intradiario y *high frequency trading* entre otros. Dependiendo de la liga que desees, y puedas, liderar obtendrás más o menos rendimientos pero siempre acorde a la infraestructura invertida. Por ejemplo, con los dos equipos que he utilizado para trabajar en el presente proyecto no podría entrar a operar en HFT aunque lo desease.

Es una materia compleja para hacerse multimillonario de la noche a la mañana pero si podemos hablar de rendimientos anuales. Con trabajo, humildad y con ayuda de Dios le animo a que lo intente.

Agradecimientos

El presente proyecto trata sobre el arte de la estimación/predicción, herramientas que nos ayudan a “adelantarnos en el tiempo” en base a una probabilidad de éxito pero la última palabra la tiene Él. Para explicar lo inexplicable hablamos de excepciones, cisnes negros, teoría del Caos, ley de Murphy. Como anécdota, yo inconformista por naturaleza y humano como el resto, alguna que otra vez cómo no he tratado de planear mi futuro pero al final ocurre como resultado otra cosa que a lo mejor resultó ser mejor de lo planeado o al revés.

Con los recursos que disponía no hubiera podido hacer ni la mitad de todo el trabajo que pretendía desarrollar. Para empezar necesitaba la muestra de datos, fuente con la que he podido procesar los resultados de todas las simulaciones. Doy gracias al proveedor1 y proveedor2 por el aporte. Las simulaciones que menciono no hubiera podido desarrollarlas si no hubiera sido por el equipamiento que dispone el Grupo de Aplicaciones de Procesado de Señal (GAPS) del Departamento de Señales, Sistemas y Radiocomunicaciones (SSR) de la E.T.S.I Telecomunicación. En concreto el equipo que lleva el Dr. Luis Hernández.

Mi tutor, el Dr. Eduardo López me ha estado ayudando con el aprendizaje de análisis cuantitativo enfocado en el *trading*. Además tuvo la brillante idea de incorporar el *trading* (proyecto ROBOTRADER) por primera vez en España en la universidad pública, principalmente dirigido a carreras de perfil técnico y le agradezco mucho que confíe en mi para construirlo juntos. Actualmente con tres nuevos fichajes que trabajan con nosotros; José Luis, Adrián y Juan.

A la E.T.S.I Telecomunicación le debo la mayoría de los conocimientos que aplico en este trabajo, el rigor que aplica su profesorado es su principal filón a la hora de reincidir en el razonamiento de los resultados y no creerse las cosas porque si.

Cómo no voy a citar a la familia, siempre ayudándome en todo lo que podían (padre, madre y hermana). Qué majos que son. Mi abuelo por ejemplo fue quien me enseño este mundo de la bolsa cuando le preguntaba durante mi estancia en el colegio, “...qué es ese gráfico, ¿esos números que significan?...”.

También me gustaría citar a mis amigos porque la paciencia que han tenido conmigo ha sido bendita, amigos de verdad sólo se cuentan con los dedos de una mano y qué importante es poder contar, para todo, con personas leales.

Una de las personas más importantes en mi vida es MG, una persona que quiero mucho y que de todo lo aprendido, gracias a ella, hasta ahora me quedo con lo que varias veces me repetía cuando adoptaba el modo “cabeza cuadrada”: ‘...Lo importante que es el cómo, más que el qué...’. Yo, siempre pretendiendo buscar la solución y la lógica del problema cuando algunas veces sólo hay que escuchar y consolar.

Lista de Palabras Clave

Sistemas de *trading* (*Algo Trading*), *quant*, *trade*, *backtest*, *equity*, *drawdown*, análisis cuantitativo, simulación, función de *fitness*, probabilidad de sobreoptimización (*overfitting*), históricos financieros, mercados financieros, *broker*, *computer science*, plataformas de procesamiento de datos, *ROBOTRADER*, Monte Carlo, algoritmo genético.

Tabla de Contenidos

Resumen.....	I
Lista de Palabras Clave.....	III
1. Introducción al trading algorítmico y su problemática	2
1.1. El <i>trading</i>	2
1.2. Discrecional VS cuantitativo	3
1.3. Motivación.....	4
1.4. Herramientas habituales	6
1.4.1. Estadística descriptiva	6
1.4.2. Series temporales	8
1.4.2.1. Método de ajuste con mínimos cuadrados.....	9
1.4.3. Métodos Monte Carlo	11
1.4.4. Optimización.....	12
1.5. Formulación del problema	18
1.6 Objetivos	19
2. Estado del arte	23
2.1. Intermediarios y mercados financieros	23
2.2. Indicadores Algorítmicos de uso frecuente	26
2.3. Sistema Algorítmico de <i>Trading</i> típico	28
2.4. Medidas de rendimiento	30
2.5. Preprocesado de los históricos financieros	35
2.6. El problema de la sobreoptimización (<i>overfitting</i>).....	46
2.6.1. Métodos de medida	46
2.6.2. Cuantificación de la sobreoptimización	49
3. Metodología.....	56
3.1. Metodología Seguida	56
3.1.1. Aspectos metodológicos.....	62
3.1.1.1. Diseño de experimentos realizados	64
3.1.1.2. Evaluación	66
3.1.2. Experimentos extras elaborados	66
3.2. Material.....	67
3.2.1. Software	67
3.2.2. Hardware	67
4. Resultados.....	69
4.1. Fase 1	69
4.2. Fase 2	74
4.2.1 Probabilidad de sobreoptimización.....	88
4.3. Fase 3	91

4.4. Experimentos número veinte y veintiuno – sin contar con fase 1.....	93
4.5. Experimento número veintidós – medidas computacionales	97
5. Conclusiones y trabajos futuros.....	98
5.1. Resumen del trabajo	98
5.2. Contribuciones del trabajo	99
5.3. Trabajos futuros	99
6. Referencias	101
ANEXO I: Características futuro E-mini S&P 500 (CME Group).....	103
ANEXO II: Informe detallado - resultados fase 1	104
ANEXO III: Informe detallado - resultados fase 2	144
ANEXO IV: Informe detallado - resultados fase 3.....	170

Lista de Figuras

Figura 1: Perfiles demandados en el sector financiero

Figura 2: procedimiento básico de un algoritmo genético

Figura 3: normalización de variables, algoritmo genético

Figura 4: distribución de probabilidades según el valor de la función fitness

Figura 5: población cruzada

Figura 6: población mutada

Figura 7: flujograma del filtro FIR-SMA en su forma canónica o tipo II

Figura 8: flujograma del filtro IIR-EMA en su forma canónica o tipo II

Figura 9: Ejemplo cruce de medias SMA y EMA

Figura 10: representación gráfica en MATLAB del retorno en el ejemplo del punto 2.4

Figura 11: ejemplo representación tipo candlestick

Figura 12: ejemplo representación tipo barra

Figura 13: ejemplo representación tipo lineal con los precios de cierre

Figura 14: captura de los apuntes de la asignatura LECM (Laboratorio de Electrónica y Comunicaciones) de una señal moduladora en el dominio del tiempo

Figura 15: captura de los apuntes de la asignatura LECM (Laboratorio de Electrónica y Comunicaciones) de una señal modulada, AM, en el dominio del tiempo

Figura 16: ejemplo señal modulada en AM con ruido en recepción en el dominio del tiempo

Figura 17: ejemplo señal moduladora, sinusoide, de frecuencia $f_m = 100$ Hz en el dominio de la frecuencia

Figura 18: Transformada de Fourier del futuro E-mini S&P 500, vencimiento de marzo 2013, timeframe 1 minuto, fuente proveedor2

Figura 19: Transformada de Fourier del futuro E-mini S&P 500, vencimiento de junio 2013, timeframe 1 minuto, fuente proveedor2

Figura 20: Transformada de Fourier del futuro E-mini S&P 500, vencimiento de septiembre 2013, timeframe 1 minuto, fuente proveedor2

Figura 21: Transformada de Fourier del futuro E-mini S&P 500, vencimiento de diciembre 2013, timeframe 1 minuto, fuente proveedor2

Figura 22: ejemplo del procedimiento Walk Forward Optimization

Figura 23: normalización del vector $\overline{\overline{R_{OOS1}}}$

Figura 24: símbolo para referenciar el procedimiento de cálculo de la probabilidad de overfitting (sobreoptimización)

Figura 25: Proceso ATurel

Figura 26: preprocessado de datos - Proceso ATurel

Figura 27: Filtrado 1 – preprocessado de datos - Proceso ATurel

Figura 28: Filtrado 2 – preprocessado de datos - Proceso ATurel

Figura 29: construcción del histórico continuo – Preprocessado de datos - Proceso ATurel

Figura 30: optimización Sistema Algorítmico de Trading + Filtros - Proceso ATurel

Figura 31: optimización Sistema Algorítmico de Trading + Filtros, en detalle - Proceso ATurel

Figura 32: PBO - Proceso ATurel

Figura 33: mejor combinación de parámetros - Proceso ATurel

Figura 34: mejor combinación de parámetros en detalle - Proceso ATurel

Figura 35: Monte Carlo - Proceso ATurel

Figura 36: Monte Carlo detallado - Proceso ATurel

Figura 37: Ratio Calmar - Proceso ATurel

Figura 38: datos mercados financieros - Proceso ATurel

Figura 39: resultados Fase 1 - Proceso ATurel

Figura 40: resultados Fase 2 - Proceso ATurel

Figura 41: resultados Fase 3 - Proceso ATurel

Figura 42: desglose de los dieciocho experimentos

Figura 43: muestra IS+OOS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

Figura 44: muestra IS + OOS, ajuste – precio cierre – E-mini S&P 500 - 15 min

Figura 45: muestra IS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

Figura 46: muestra IS, ajuste – precio cierre – E-mini S&P 500 - 15 min

Figura 47: muestra OOS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

Figura 48: muestra OOS, ajuste – precio cierre – E-mini S&P 500 - 15 min

Figura 49: consumo de RAM inicial – Sistema

Figura 50: consumo de RAM inicial con MATLAB arrancado – Sistema

Figura 51: consumo RAM – Experimento 1, 2 y 3

Figura 52: carga del sistema – 6 cores – Experimento 1, 2 y 3

Figura 53: carga media del sistema – Experimento 1, 2 y 3

Figura 54: nube de puntos del resultado de la función *fitness* IS – Experimento 1, 2 y 3

Figura 55: nube de puntos del resultado de la función *fitness* OOS – Experimento 1, 2 y 3

Figura 56: histograma de los valores de la función *fitness* IS – Experimento 1, 2 y 3

Figura 57: *equity* IS – Experimento 1,2 y 3

Figura 58: *equity* OOS – Experimento 1,2 y 3

Figura 59: *equity* IS + OOS – Experimento 1,2 y 3

Figura 60: consumo RAM – Experimento 10, 11 y 12

Figura 61: carga del sistema – 6 cores – Experimento 10, 11 y 12

Figura 62: carga media del sistema – Experimento 10, 11 y 12

Figura 63: nube de puntos del resultado de la función *fitness* IS – Experimento 10, 11 y 12

Figura 64: nube de puntos del resultado de la función *fitness* OOS – Experimento 10, 11 y 12

Figura 65: histograma de los valores de la función *fitness* IS – Experimento 10, 11 y 12

Figura 66: *equity* IS – Experimento 10, 11 y 12

Figura 67: *equity* OOS – Experimento 10, 11 y 12

Figura 68: *equity* IS + OOS – Experimento 10, 11 y 12

Figura 69: consumo *RAM* – Experimento 13, 14 y 15

Figura 70: carga del sistema – 6 cores – Experimento 13, 14 y 15

Figura 71: carga media del sistema – Experimento 13, 14 y 15

Figura 72: nube de puntos del resultado de la función *fitness* IS – Experimento 13, 14 y 15

Figura 73: nube de puntos del resultado de la función *fitness* OOS – Experimento 13, 14 y 15

Figura 74: histograma de los valores de la función *fitness* IS – Experimento 13, 14 y 15

Figura 75: *equity* IS – Experimento 13, 14 y 15

Figura 76: *equity* OOS – Experimento 13, 14 y 15

Figura 77: *equity* IS + OOS – Experimento 13, 14 y 15

Figura 78: retornos IS – rango fijo de parámetros – 10404 combinaciones

Figura 79: retornos OOS – rango fijo de parámetros – 10404 combinaciones

Figura 80: retornos IS – rango fijo de parámetros – 5202 combinaciones

Figura 81: retornos OOS – rango fijo de parámetros – 5202 combinaciones

Figura 82: retornos IS – rango fijo de parámetros – 2601 combinaciones

Figura 83: retornos OOS – rango fijo de parámetros – 2601 combinaciones

Figura 84: *equity* IS – sin fase 1 – proveedor1

Figura 85: *equity* OOS – sin fase 1 – proveedor1

Figura 86: *equity* IS + OOS – sin fase 1 – proveedor1

Figura 87: *equity* IS – sin fase 1 – proveedor2

Figura 88: *equity OOS – sin fase 1 – proveedor2*

Figura 89: *equity OOS – sin fase 1 – proveedor2*

Figura 90: carga del sistema – no paralelo – Experimento 13, 14 y 15

Figura 91: carga media del sistema – no paralelo – Experimento 13, 14 y 15

Figura 92: carga del sistema – 6 cores – Experimento 13, 14 y 15

Figura 93: carga media del sistema – Experimento 13, 14 y 15

Figura 94: Muestra IS+OOS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

Figura 95: Muestra IS + OOS, ajuste – precio cierre – E-mini S&P 500 - 15 min

Figura 96: Muestra IS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

Figura 97: Muestra IS, ajuste – precio cierre – E-mini S&P 500 - 15 min

Figura 98: Muestra OOS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

Figura 99: Muestra OOS, ajuste – precio cierre – E-mini S&P 500 - 15 min

Figura 100: Período 1 de un tramo lateral

Figura 101: Período 2 de un tramo lateral

Figura 102: Período 3 de un tramo lateral

Figura 103: Período 4 de un tramo lateral

Figura 104: Período 5 de un tramo lateral

Figura 105: Período 6 de un tramo lateral

Figura 106: Período 7 de un tramo lateral

Figura 107: Período 8 de un tramo lateral

Figura 108: Período 9 de un tramo lateral

Figura 109: Período 10 de un tramo lateral

Figura 110: Consumo de RAM inicial – Sistema

Figura 111: Consumo de RAM inicial con MATLAB arrancado – Sistema

Figura 112: Consumo RAM – Experimento 1, 2 y 3

Figura 113: Carga del sistema – 6 cores – Experimento 1, 2 y 3

Figura 114: Carga media del sistema – Experimento 1, 2 y 3

Figura 115: Nube de puntos del resultado de la función *fitness* IS – Experimento 1, 2 y 3

Figura 116: Nube de puntos del resultado de la función *fitness* OOS – Experimento 1, 2 y 3

Figura 117: Histograma de los valores de la función *fitness* IS – Experimento 1, 2 y 3

Figura 118: *Equity* IS – Experimento 1,2 y 3

Figura 119: *Equity* OOS – Experimento 1,2 y 3

Figura 120: *Equity* IS + OOS – Experimento 1,2 y 3

Figura 121: Consumo RAM – Experimento 4, 5 y 6

Figura 122: Carga del sistema – 6 cores – Experimento 4, 5 y 6

Figura 123: Carga media del sistema – Experimento 4, 5 y 6

Figura 124: Nube de puntos del resultado de la función *fitness* IS – Experimento 4, 5 y 6

Figura 125: Nube de puntos del resultado de la función *fitness* OOS – Experimento 4, 5 y 6

Figura 126: Histograma de los valores de la función *fitness* IS – Experimento 4, 5 y 6

Figura 127: *Equity* IS – Experimento 4, 5 y 6

Figura 128: *Equity* OOS – Experimento 4, 5 y 6

Figura 129: *Equity* IS + OOS – Experimento 4, 5 y 6

Figura 130: Consumo RAM – Experimento 7, 8 y 9

Figura 131: Carga del sistema – 6 cores – Experimento 7, 8 y 9

Figura 132: Carga media del sistema – Experimento 7, 8 y 9

Figura 133: Nube de puntos del resultado de la función *fitness* IS – Experimento 7, 8 y 9

Figura 134: Nube de puntos del resultado de la función *fitness* OOS – Experimento 7, 8 y 9

Figura 135: Histograma de los valores de la función *fitness* IS – Experimento 7, 8 y 9

Figura 136: *Equity* IS – Experimento 7, 8 y 9

Figura 137: *Equity* OOS – Experimento 7, 8 y 9

Figura 138: *Equity* IS + OOS – Experimento 7, 8 y 9

Figura 139: Consumo *RAM* – Experimento 10, 11 y 12

Figura 140: Carga del sistema – 6 cores – Experimento 10, 11 y 12

Figura 141: Carga media del sistema – Experimento 10, 11 y 12

Figura 142: Nube de puntos del resultado de la función *fitness* IS – Experimento 10, 11 y 12

Figura 143: Nube de puntos del resultado de la función *fitness* OOS – Experimento 10, 11 y 12

Figura 144: Histograma de los valores de la función *fitness* IS – Experimento 10, 11 y 12

Figura 145: *Equity* IS – Experimento 10, 11 y 12

Figura 146: *Equity* OOS – Experimento 10, 11 y 12

Figura 147: *Equity* IS + OOS – Experimento 10, 11 y 12

Figura 148: Consumo *RAM* – Experimento 13, 14 y 15

Figura 149: Carga del sistema – 6 cores – Experimento 13, 14 y 15

Figura 150: Carga media del sistema – Experimento 13, 14 y 15

Figura 151: Nube de puntos del resultado de la función *fitness* IS – Experimento 13, 14 y 15

Figura 152: Nube de puntos del resultado de la función *fitness* OOS – Experimento 13, 14 y 15

Figura 153: Histograma de los valores de la función *fitness* IS – Experimento 13, 14 y 15

Figura 154: *Equity* IS – Experimento 13, 14 y 15

Figura 155: *Equity* OOS – Experimento 13, 14 y 15

Figura 156: *Equity* IS + OOS – Experimento 13, 14 y 15

Figura 157: Consumo RAM – Experimento 16, 17 y 18

Figura 158: Carga del sistema – 6 cores – Experimento 16, 17 y 18

Figura 159: Carga media del sistema – Experimento 16, 17 y 18

Figura 160: Nube de puntos del resultado de la función *fitness* IS – Experimento 16, 17 y 18

Figura 161: Nube de puntos del resultado de la función *fitness* OOS – Experimento 16, 17 y 18

Figura 162: Histograma de los valores de la función *fitness* IS – Experimento 16, 17 y 18

Figura 163: *Equity* IS – Experimento 16, 17 y 18

Figura 164: *Equity* OOS – Experimento 16, 17 y 18

Figura 165: *Equity* IS + OOS – Experimento 16, 17 y 18

Figura 166: Nube de puntos del resultado de la función *fitness* OOS – Ejemplo de la evolución de sobreoptimización

Lista de Tablas

Tabla 1: código vencimientos futuro de enero a agosto

Tabla 2: código vencimientos futuro de septiembre a diciembre

Tabla 3: tramos laterales, histórico continuo filtrado y construido

Tabla 4: mejor combinación de parámetros – Experimento 1, 2 y 3

Tabla 5: retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 1, 2 y 3

Tabla 6: *trades* en tramo alcista | bajista | lateral – Experimento 1, 2 y 3

Tabla 7: mejor combinación de parámetros – Experimento 10, 11 y 12

Tabla 8: retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 10, 11 y 12

Tabla 9: *trades* en tramo alcista | bajista | lateral – Experimento 10, 11 y 12

Tabla 10: mejor combinación de parámetros – Experimento 13, 14 y 15

Tabla 11: retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 13, 14 y 15

Tabla 12: *trades* en tramo alcista | bajista | lateral – Experimento 13, 14 y 15

Tabla 13: retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 1, 2 y 3

Tabla 14: retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 10, 11 y 12

Tabla 15: retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 13, 14 y 15

Tabla 16: mejor combinación de parámetros

Tabla 17: datos IS – sin fase 1 – proveedor1

Tabla 18: datos OOS – sin fase 1 – proveedor1

Tabla 19: datos IS – sin fase 1 – proveedor2

Tabla 20: datos OOS – sin fase 1 – proveedor2

Tabla 21: tiempos de ejecución

Tabla 22: Filtrado 1, vencimiento ESH10

Tabla 23: Filtrado 1, vencimiento ESH11

Tabla 24: Filtrado 1, vencimiento ESH12

Tabla 25: Filtrado 1, vencimiento ESH13

Tabla 26: Filtrado 1, vencimiento ESM10

Tabla 27: Filtrado 1, vencimiento ESM11

Tabla 28: Filtrado 1, vencimiento ESM12

Tabla 29: Filtrado 1, vencimiento ESM13

Tabla 30: Filtrado 1, vencimiento ESU10

Tabla 31: Filtrado 1, vencimiento ESU11

Tabla 32: Filtrado 1, vencimiento ESU12

Tabla 33: Filtrado 1, vencimiento ESU13

Tabla 34: Filtrado 1, vencimiento ESZ10

Tabla 35: Filtrado 1, vencimiento ESZ11

Tabla 36: Filtrado 1, vencimiento ESZ12

Tabla 37: Filtrado 1, vencimiento ESZ13

Tabla 38: Filtrado 2, vencimiento ESH10

Tabla 39: Filtrado 2, vencimiento ESH11

Tabla 40: Filtrado 2, vencimiento ESH12

Tabla 41: Filtrado 2, vencimiento ESH13

Tabla 42: Filtrado 2, vencimiento ESM10

Tabla 43: Filtrado 2, vencimiento ESM11

Tabla 44: Filtrado 2, vencimiento ESM12

Tabla 45: Filtrado 2, vencimiento ESM13

Tabla 46: Filtrado 2, vencimiento ESU10

Tabla 47: Filtrado 2, vencimiento ESU11

Tabla 48: Filtrado 2, vencimiento ESU12

Tabla 49: Filtrado 2, vencimiento ESU13

Tabla 50: Filtrado 2, vencimiento ESZ10

Tabla 51: Filtrado 2, vencimiento ESZ11

Tabla 52: Filtrado 2, vencimiento ESZ12

Tabla 53: Filtrado 2, vencimiento ESZ13

Tabla 54: Máxima negociación – proveedor1

Tabla 55: Máxima negociación – proveedor2

Tabla 56: rollover entre ESH10 y ESM10

Tabla 57: rollover entre ESM10 y ESU10

Tabla 58: rollover entre ESU10 y ESZ10

Tabla 59: rollover entre ESZ10 y ESH11

Tabla 60: rollover entre ESH11 y ESM11

Tabla 61: rollover entre ESM11 y ESU11

Tabla 62: rollover entre ESU11 y ESZ11

Tabla 63: rollover entre ESZ11 y ESH12

Tabla 64: rollover entre ESH12 y ESM12

Tabla 65: rollover entre ESM12 y ESU12

Tabla 66: rollover entre ESU12 y ESZ12

Tabla 67: rollover entre ESZ12 y ESH13

Tabla 68: rollover entre ESH13 y ESM13

Tabla 69: rollover entre ESM13 y ESU13

Tabla 70: rollover entre ESU13 y ESZ13

Tabla 71: Período 1 de un tramo lateral

Tabla 72: Período 2 de un tramo lateral

Tabla 73: Período 3 de un tramo lateral

Tabla 74: Período 4 de un tramo lateral

Tabla 75: Período 5 de un tramo lateral

Tabla 76: Período 6 de un tramo lateral

Tabla 77: Período 7 de un tramo lateral

Tabla 78: Período 8 de un tramo lateral

Tabla 79: Período 9 de un tramo lateral

Tabla 80: Período 10 de un tramo lateral

Tabla 81: Mejor combinación de parámetros – Experimento 1, 2 y 3

Tabla 82: Retorno, máximo drawdown y función fitness de la mejor combinación de parámetros en IS y OOS – Experimento 1, 2 y 3

Tabla 83: Trades en tramo alcista | bajista | lateral – Experimento 1, 2 y 3

Tabla 84: Mejor combinación de parámetros – Experimento 4, 5 y 6

Tabla 85: Retorno, máximo drawdown y función fitness de la mejor combinación de parámetros en IS y OOS – Experimento 4, 5 y 6

Tabla 86: Trades en tramo alcista | bajista | lateral – Experimento 4, 5 y 6

Tabla 87: Mejor combinación de parámetros – Experimento 7, 8 y 9

Tabla 88: Retorno, máximo drawdown y función fitness de la mejor combinación de parámetros en IS y OOS – Experimento 7, 8 y 9

Tabla 89: Trades en tramo alcista | bajista | lateral – Experimento 7, 8 y 9

Tabla 90: Mejor combinación de parámetros – Experimento 10, 11 y 12

Tabla 91: Retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 10, 11 y 12

Tabla 92: *Trades* en tramo alcista | bajista | lateral – Experimento 10, 11 y 12

Tabla 93: Mejor combinación de parámetros – Experimento 13, 14 y 15

Tabla 94: Retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 13, 14 y 15

Tabla 95: *Trades* en tramo alcista | bajista | lateral – Experimento 13, 14 y 15

Tabla 96: Mejor combinación de parámetros – Experimento 16, 17 y 18

Tabla 97: Retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 16, 17 y 18

Tabla 98: *Trades* en tramo alcista | bajista | lateral – Experimento 16, 17 y 18

Tabla 99: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 1, 2 y 3

Tabla 100: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 4, 5 y 6

Tabla 101: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 7, 8 y 9

Tabla 102: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 10, 11 y 12

Tabla 103: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 13, 14 y 15

Tabla 104: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 16, 17 y 18

1. Introducción al trading algorítmico y su problemática

“When a beginner wins he feels brilliant and invincible, then he takes wild risk and loses everything.”

Alexander Elder

Alexander Elder autor de varios libros sobre *trading*, libros que yo personalmente no he leído pero que resume en una frase la verdad de lo que les pasa a la mayoría de personas que empezamos en un oficio como lo es el *trading* “Cuando un principiante gana se siente brillante e invencible, después adquiere más riesgo (riesgo salvaje) y lo pierde todo”.

Vamos a ver a continuación procedimientos de estudio y que en particular cito las palabras **análisis** y **riesgo**, siendo éstas determinantes para no perderlo todo.

1.1. El *trading*

La negociación bursátil comenzó en el siglo XVI con los pagarés y bonos. Gracias a la implantación del mercado de futuros, los productores de materias primas (como lo fue en su día la moda del cultivo de tulipán) podían asegurar su venta a fin de vencimiento (en el caso del tulipán al final de la época de floración). Después llegaron los especuladores (*traders*) alrededor del siglo XVII valorando que se podía ganar dinero con la variación del precio y que sin una volatilidad alarmante, a los productores les beneficiaba que aumentara el volumen de negociación (liquidez) y lógicamente el aumento del precio para su venta.

El *trading* es la operativa de compraventa de activos (productos financieros) por cuenta propia (a través de un intermediario financiero, generalmente un *broker*, para acceder a mercado), con expectativa de obtención de beneficios a causa de la modificación del precio de un mercado, la operativa más popularizada de tipo *swing* (intradiario y/o diario-semanal) y de tipo *scalping* (intradiario y como más innovador donde la tecnología juega el papel importante, *High Frequency Trading*). Dicha compraventa se puede llevar a cabo siguiendo un conjunto de reglas y/o instrucciones implementadas en un modelo llamado Sistema Algorítmico de *Trading* (adaptado lógicamente, entre otros, por el tipo de operativa) que pueden ejecutarse de forma automatizada. **La operativa en productos financieros se ha automatizado por completo** (Más del 75% del volumen de negociación del CME Group provienen de operaciones realizadas electrónicamente en su plataforma de negociación) y permite la obtención de beneficios para aquellos sistemas que son capaces de estimar correctamente el sentido de las variaciones del precio. Las variaciones son función de innumerables variables económicas, políticas, sociales, temporales y psicológicas. Además se pueden medir dichas variaciones mediante el análisis cuantitativo del histórico bursátil (*historical data feed*), a continuación seleccionar una muestra del histórico para la optimización (muestra *In Sample*) y la muestra para las pruebas *backtest* (muestra *Out Of Sample*).

Este procedimiento de elegir una muestra para optimizar (entrenar el modelo) y otra para probar es utilizado desde hace años por investigadores en robótica, predicción de condiciones atmosféricas y sistemas de comunicaciones entre otros.

Finalmente postprocesado de los resultados con métodos estadísticos/probabilísticos para analizar tanto la robustez como el riesgo, siendo ésta una variable clave que debe conocer el inversor, supone la inversión de una determinada cartera (*portfolio*) de productos financieros y/o Sistemas Algorítmicos de *Trading*.

1.2. Discrecional VS cuantitativo

En la época citada en el punto anterior no había ordenadores para realizar estudios cuantitativos con grandes volúmenes de datos para analizar. La operativa, sea discrecional o cuantitativa puede ser más o menos arriesgada y de manera, o no, sistemática, pero bien es cierto que un análisis cuantitativo de los datos, sean los retornos de una empresa, los síntomas de una enfermedad o el comportamiento humano entre otros, va a resultar mucho más beneficioso para entrenar objetivamente el modelo a ejecutar. A lo largo de la historia *traders* con ausencia de conocimientos en ciencias para verificar de manera objetiva por qué el precio sube o baja en términos estadísticos/probabilísticos ha ocasionado grandes burbujas (Tulipán, Mississippi, Mar del Sur, Ferrocarril británico e inmobiliario entre otros) que finalmente explotaban causando las llamadas ‘crisis’. A día de hoy seguimos sufriendo el fenómeno ‘crisis’ a pesar de la supuesta experiencia en la materia en el transcurso de los últimos cientos de años y que se volverán a repetir hasta que no aprendamos de nuestros errores o porque así lo deseé algún *lobby* concreto. Un procedimiento no sistemático y menos cuantitativo es aquel que de forma cualitativa en su mayoría tienen como fundamento ‘...compro porque creo que el precio va a subir, tiene buena pinta...hay que vender porque me lo ha dicho un contacto...’. Cito como anécdota que el *trader* discrecional Nick Leeson hundió, el *Barings Bank*, insolvente en 1995, el banco privado más antiguo de Gran Bretaña (fundado en 1762) y comprado por 1 libra por el banco y compañía de seguros holandés ING. Se puede ver un resumen de lo que ocurrió en la película *El Fraude*. Por cierto, uno de sus fundamentos fue “Si me ha salido bien una vez, puede que haya descubierto la forma de hacerme rico”.

Pasamos directamente al final del siglo XX con la incorporación del ordenador personal como herramienta y la evolución de internet. La innovación tecnológica (en términos de servicios, protocolos de comunicación) de los últimos años ha permitido la automatización de los mercados financieros (mercados electrónicos) y la forma de operativa. Aunque el 75% de las operaciones se realizan automáticamente en la plataforma electrónica, como citábamos en el punto anterior, no significa que éstas se fundamenten con procedimientos cuantitativos algorítmicos. Sin ir más lejos, el proyecto *ROBOTRADER* (empezó en 2009) es el primer proyecto en España que cubre la integración de *Computer Science* y Mercados Financieros. Bien es sabido los que hemos estudiado carreras técnicas, como en mi caso Ingeniería de Telecomunicación, donde tenemos la suerte de cursar asignaturas como ‘Arquitectura de Ordenadores’.

Las ventajas que ofrecen los estudios con un ordenador (o varios en paralelo) frente a un papel y lápiz, entre otros el poder realizar operaciones a una velocidad de varios GHz y con mayor o menor número de núcleos (*cores*), contando el número de hebras por núcleo, para poder paralelizar.

Conozco a pocas personas que puedan almacenar en su cabeza ficheros de GB y que además puedan realizar cálculos en paralelo a velocidades de GHz. Otra de las ventajas de los análisis cuantitativos radica en la identificación de patrones y como cabe de esperar también contempla inconvenientes. El primer principal inconveniente, el desconocimiento del procedimiento que se está analizando y por ende creerse los resultados dados en vez de razonarlos (durante mi carrera universitaria, el rigor del profesorado ETSIT-UPM ha jugado un papel magnífico en cuanto a razonar todo resultado) y como segundo principal inconveniente que estamos utilizando máquinas y programas informáticos creados por personas, por lo tanto no son perfectos (los profesionales de la ciberseguridad y ciberdefensa saben a lo que me refiero). Como anécdota tenemos el *Flash Crash de 2010* donde los sistemas que se nutrían, entre otros, por información de redes sociales tomaron datos erróneos (suplantación de identidad) y ésto provocó que el índice Dow Jones se desplomase 1014.14 puntos y que a los pocos minutos recuperase 998.5 puntos. Este es uno de los ejemplos que les preocupa a los gobiernos porque básicamente no saben de qué trata este tema tecnológico y si aun no sabiéndolo lo tratan para creerse que lo tienen controlado (implantación de la *tasa Tobin*) entonces distorsionan la oferta-demanda trayendo consigo posibles malas consecuencias (liquidez por ejemplo). A mi juicio no se puede operar en contra del mercado, además no se puede y no se debe controlar el mercado porque la oferta y la demanda, manda.

1.3. Motivación

Para los integrantes de un mercado, puede ser interesante utilizar programas para realizar un análisis cuantitativo y sistemático de la posible evolución de los precios de los productos así como el riesgo de invertir en los mismos. Dichos análisis presentan la ventaja de aportar resultados calculados únicamente según los parámetros y funciones que se deseen, aislados de la componente psicológica del inversor. La parte psicológica del *trading*, aunque no vayamos a entrar para estudiarla porque conseguimos extraerla gracias al análisis cuantitativo y la operativa con Sistemas Algorítmicos de *Trading*, siempre queda condicionado en un porcentaje, aun siendo pequeño, como es por ejemplo la variable de máximo tiempo en *drawdown*. Es decir, máximo tiempo que el sistema se encuentra en pérdidas. Si por ejemplo tenemos un sistema que realiza cinco operaciones de media al mes y lleva tres meses con pérdidas, el *trader* debería estar seguro de decidir si es según lo esperado (estimaciones del estudio) o si en cambio resulta que el sistema ya no funciona (concepto de ‘Tiempo de Vida’ de un sistema).

Bien se habrá dado cuenta el lector, en lo que lleva de lectura de este PFC, mi defensa del análisis cuantitativo, pero no sólo lo defiendo yo personalmente por tener la suerte de haber estudiado una carrera técnica, ingeniería, sino que las empresas que tienen relación con el sector financiero como son los bancos, *brokers*, desarrolladores de sistemas y/o plataformas, reguladores de mercado y CTAs entre otros están demandando cada vez más personal de perfil técnico con conocimientos cuantitativos, que si además dicha persona conoce el arte de la programación con conocimientos en tecnología (*big data* y redes entre otros) y algo de mercados financieros cumple con la integración perfecta que desean estas empresas. Se le está dando el término de *quant* al nuevo perfil que necesitan empresas financieras junto con los ya conocidos *traders*. Esto es sabido gracias al proyecto ROBOTRADER como citaba en el punto 1.2, como director el Dr. Eduardo López, el cual me ha dado la oportunidad de construir con él dicho proyecto, para comprobar la amplia colaboración de empresas por la cantidad de herramientas de pago (licencias de plataformas de trading y cuentas de datos en tiempo real entre otros) que se reparten totalmente de forma gratuita a los estudiantes para entrenar y competir.

Antes de concluir, englobamos los principales roles demandados en la industria financiera:

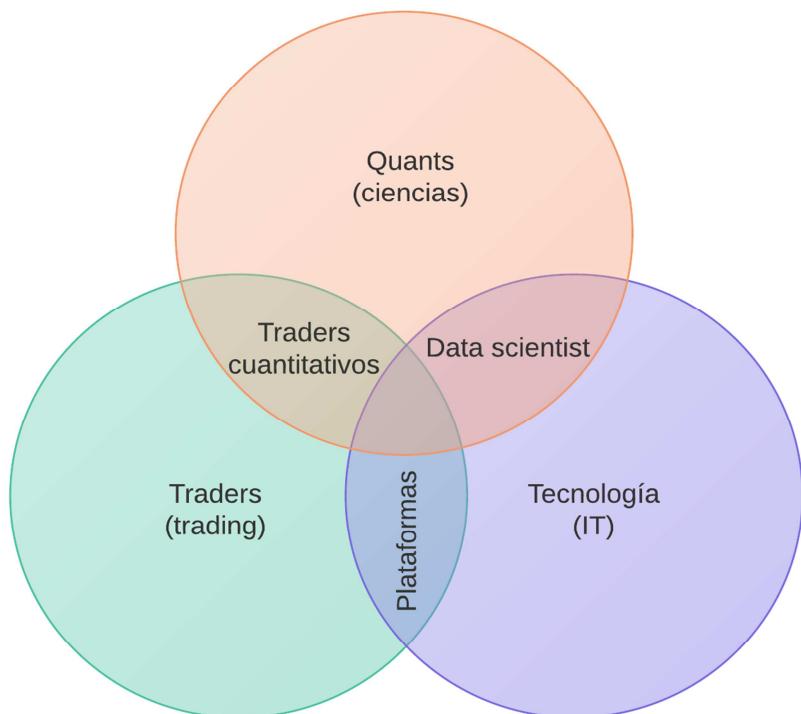


Figura 1: Perfiles demandados en el sector financiero

Hablamos de trabajo en equipo con varios perfiles como se puede ver en el anterior diagrama de Venn. De forma muy resumida podríamos decir que los *traders* aportan la idea, los *quants* formalizan el modelo de la idea y los de tecnología mantienen la conexión con el mercado para que el modelo continúe su ejecución normal.

1.4. Herramientas habituales

Las herramientas y procedimientos de estudio más habituales que se aplican en la ciencia se están demandando ahora más que nunca en el campo del *trading*. Las matemáticas es la ciencia y la base para llegar a un razonamiento lógico, no podemos imaginarnos vivir el día a día sin nombrarlas (la tan famosa, ahora, prima de riesgo ha bajado a los 120 puntos básicos, este producto es un 2% más barato que este otro, el mercado está correlacionado con las decisiones humanas).

¿Quiénes son los mejores programadores? Los matemáticos, ¿Quiénes son los mejores formulando proposiciones? Los matemáticos. El resto, me incluyo, aunque no llegamos al nivel de rendimiento que un matemático pueda dar podemos estar tranquilos porque dicen (lo defiendo) que cuanto más simple se modele el problema mejor y como además somos rigurosos con el procedimiento de estudio, obtendremos muy buenos resultados. Si no que se lo digan a los ingenieros aeronáuticos que han conseguido que vuele un avión o los ingenieros de telecomunicación que año tras año consiguen mejorar el BER (Bit Error Rate) de un canal de fibra óptica. Los llamados *quants* que mencionábamos en el punto anterior, a lo mejor no saben nada del análisis técnico que un *trader* si conoce, en cambio el *quant* es una persona cualificada para el análisis de datos para la identificación de patrones y ésto último es lo que comentaba que se demanda cada vez más. Si encima el *quant* es una persona íntegra y demuestra otros conocimientos como los de un *trader* o de tecnología, entonces todo suma, lógicamente.

1.4.1. Estadística descriptiva

Ciencia que analiza los datos (discretos o continuos), de una población o muestra, de variables cuantitativas (precios, *drawdown*, *trades* y ratios entre otros) para poder extraer, a través de valores numéricos como también de gráficas, conclusiones de variables tanto cualitativas (comportamiento del mercado, día alcista o bajista o lateral entre otros) como cuantitativas (beneficio medio y dispersión respecto a la media entre otros).

Una variable estadística es una cualidad susceptible de ser medida y que toma intrínsecamente valores numéricos (uno sólo o distribuidos en un intervalo). Generalmente se toma como representante del intervalo el valor central, ‘Marca de Clase’). Nos referimos al término ‘Descriptiva’ porque para la correcta identificación de cada variable se le asigna una etiqueta descriptiva.

La población es el conjunto total de los datos, dado un problema con variables para analizar.

Una muestra es un subconjunto de los datos de la población, dado un problema con variables para analizar.

Dada una distribución de frecuencias $(x_i, n_i) | i = 1, 2, \dots, k$

x_i == valores que toma una determinada variable estadística.

n_i == número de veces que aparece el valor x_i entre los datos (frecuencia absoluta).

N == número total de datos.

Por si nos resulta más cómodo otro tipo de representación, podemos utilizar la frecuencia relativa;

$$f_i = \frac{n_i}{N}$$

Durante el estudio del presente PFC utilizaremos, entre la gran variedad de variables de estudio existentes en la estadística descriptiva, lo siguiente:

- Representaciones gráficas
 - No agrupadas en intervalos
 - Diagrama de barras: se puede representar tanto la frecuencia absoluta como la frecuencia relativa.
 - Agrupadas en intervalos
 - Histograma: el área encerrada en cada intervalo es equivalente a la frecuencia absoluta. Puede ocurrir que los intervalos sean de distinta amplitud.
- Medidas de tendencia central (muestral):
 - Media aritmética
$$\bar{x} = \frac{1}{N} * \sum_{i=1}^k x_i$$
 - Mediana (Me): útil respecto a la media aritmética cuando la distribución presenta unos pocos valores extremos (llamadas generalmente 'casualidades estadísticas'). Se ordena el vector de los datos y se calcula la mediana, ésta deja el mismo número de datos por la izquierda que por la derecha. Si la mediana cae entre dos valores se aplica generalmente la media aritmética.

- Medidas de dispersión (muestral):
 - Desviación respecto a promedios (generalmente la media y mediana)
 - Desviación media con respecto a la media (más conocida como desviación típica)
$$S = \frac{1}{N-1} * \sum_{i=1}^k |x_i - \bar{x}|$$
 - Desviación media con respecto a la mediana

$$S_{Me} = \frac{1}{N-1} * \sum_{i=1}^k |x_i - Me|$$

- Varianza

$$S^2 = \frac{1}{N-1} * \sum_{i=1}^k |x_i - \bar{x}|^2$$

Utilizaremos la estadística descriptiva para el postprocesado de datos, es decir, después de cada simulación a realizar calcularemos las medidas citadas medidas y representaciones gráficas también citadas para poder dar conclusiones.

1.4.2. Series temporales

Conjunto de datos numéricos referidos a instantes temporales (años, meses, días, horas, minutos, segundos). Para cada valor temporal (el desarrollo de este PFC serán minutos) tenemos un valor, en este caso el precio de cierre, de lo que estamos midiendo en la serie. Se trata de una señal en tiempo discreto.

Partiendo de estos datos establecemos un modelo. Utilizando las características de este modelo podemos hacer lo siguiente:

- Interpolación: En el caso de que se haya perdido algún valor de la serie, con las características del modelo podemos aproximar el valor. Lo utilizaremos con una curva de regresión en el preprocesado de datos (comparando varias curvas para saber cuál se ajusta mejor a la serie) y con una recta de regresión (para ajustar la *equity* y controlar el riesgo) en el postprocesado de datos para elegir, entre otros procedimientos, la mejor combinación de parámetros.

- Extrapolación/predicción: Tratamos de predecir los valores futuros con las características del modelo. Estas predicciones se emplearán para la toma de decisiones. El presente Proyecto Fin de Carrera está estudiado y elaborado para dar resultados y conclusiones de análisis de sistemas, no para elaborar un sistema (conllevaría a la realización de otro Proyecto Fin de Carrera o incluso Tesis Doctoral). El sistema empleado se explicará más adelante que es tradicionalmente típico y que gracias a lo explicado en este trabajo servirá como ejemplo de estudio de análisis.

1.4.2.1. Método de ajuste con mínimos cuadrados

Una de sus aplicaciones (la que necesitamos para el trabajo) es al ajuste de un conjunto de datos con funciones polinómicas.

Sea $x[n], n=1,2,\dots, N$ un conjunto de observaciones correspondientes, por ejemplo, a N minutos consecutivos de un histórico financiero. Se desea ajustar a $x[n]$ el modelo polinómico:

$$x[n] = a_0 + a_1n + \dots + a_ln^l$$

Si fuera posible realizar exactamente el ajuste, debería resolverse el siguiente sistema de ecuaciones:

$$\left. \begin{array}{l} x[1] = a_0 + a_11^1 + a_21^2 + \dots + a_l1^l \\ x[2] = a_0 + a_12^1 + a_22^2 + \dots + a_l2^l \\ \vdots \\ x[N] = a_0 + a_1N^1 + a_2N^2 + \dots + a_lN^l \end{array} \right\}$$

Pero debe notarse que se trata de un sistema de ecuaciones con más ecuaciones que incógnitas, ya que el orden del polinomio de ajuste (l) será en general bajo, mientras que el número de puntos disponibles de señal (N) puede ser alto. Por tanto, en general, salvo que los puntos $x[n]$ ya se encuentren en un polinomio, no será posible cumplir las ecuaciones del sistema anterior. Lo modificaremos entonces así:

$$\left. \begin{array}{l} x[1] \approx a_0 + a_11^1 + a_21^2 + \dots + a_l1^l = s[1] \\ x[2] \approx a_0 + a_12^1 + a_22^2 + \dots + a_l2^l = s[2] \\ \vdots \\ x[N] \approx a_0 + a_1N^1 + a_2N^2 + \dots + a_lN^l = s[N] \end{array} \right\}$$

El problema se puede plantear ahora para encontrar el conjunto de valores a_i , que definen el polinomio de ajuste, de forma que los valores de señal $x[n], n=1,2,\dots, N$ y del polinomio de ajuste $s[n], n=1,2,\dots, N$ "se parezcan" lo más posible. Esto exige definir de forma precisa y cuantitativa el concepto de "parecido".

Concretamente, buscaremos el conjunto de coeficientes a_i de forma que se minimice el error cuadrático (minimizar la suma de los cuadrados de las distancias verticales entre los valores de los datos y los de la regresión estimada, el polinomio):

$$J(a_0, a_1, \dots, a_l) = \sum_{i=1}^N (x[i] - s[i])^2$$

Si definimos los vectores y matrices:

$$\underline{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_l \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[N] \end{bmatrix}, \quad \underline{s} = \begin{bmatrix} s[1] \\ s[2] \\ \vdots \\ s[N] \end{bmatrix}, \quad \underline{\underline{H}} = \begin{bmatrix} 1 & 1^1 & \cdots & 1^l \\ 1 & 2^1 & \cdots & 2^l \\ \vdots & \vdots & \ddots & \vdots \\ 1 & N^1 & \cdots & N^l \end{bmatrix}$$

el vector \underline{s} se puede expresar como

$$\underline{s} = \underline{\underline{H}} \underline{a}$$

y el error cuadrático J como el producto escalar

$$J(\underline{a}) = (\underline{x} - \underline{s}(\underline{a}))^T (\underline{x} - \underline{s}(\underline{a}))$$

donde se ha puesto explícitamente la dependencia de \underline{s} con el vector de parámetros \underline{a} .

El problema es por tanto obtener el vector \underline{a} que minimice $J(\underline{a})$. Operando:

$$(\underline{x} - \underline{s})^T (\underline{x} - \underline{s}) = (\underline{x} - \underline{\underline{H}} \underline{a})^T (\underline{x} - \underline{\underline{H}} \underline{a}) = \underline{x}^T \underline{x} - 2\underline{a}^T \underline{\underline{H}}^T \underline{x} + \underline{a}^T \underline{\underline{H}}^T \underline{\underline{H}} \underline{a}$$

donde el superíndice T indica traspuesta y no hay que olvidar que $J(\underline{a})$ es un escalar.

Para obtener el mínimo se aplica:

$$\frac{d(J(\underline{a}))}{d\underline{a}} = 0$$

Teniendo en cuenta que:

$$\frac{d(\underline{a}^T \underline{\underline{H}}^T)}{d\underline{a}} = \underline{\underline{H}} \quad \frac{d(\underline{a}^T \underline{\underline{H}}^T \underline{\underline{H}} \underline{a})}{d\underline{a}} = 2\underline{\underline{H}}^T \underline{\underline{H}} \underline{a}$$

y que el vector de datos \underline{x} no depende de \underline{a} , se obtiene

$$\frac{d(J(\underline{a}))}{d\underline{a}} = -2\underline{\underline{H}}\underline{x} + 2\underline{\underline{H}}^T \underline{\underline{H}}\underline{a} = 0$$

y por tanto

$$\underline{a} = (\underline{\underline{H}}^T \underline{\underline{H}})^{-1} \underline{\underline{H}}^T \underline{x}$$

Un polinomio (comparando distintos grados) será empleado para el preprocesado de datos financieros, para ofrecer tanto una estimación de posibles valores de la serie que se hayan perdido (*gaps*) como para marcar la tendencia de la serie para dividir los resultados de la simulación en alcista | bajista | lateral (hallando la pendiente de la recta tangente en un punto dado). Por otro lado, la recta de regresión (polinómica de orden 1) junto con el coeficiente de determinación o cuadrado del coeficiente de correlación lineal de Pearson* nos va a ayudar con el postprocesado de datos para seleccionar unos parámetros u otros a razón de los resultados (*equities*) de la simulación OOS.

1.4.3. Métodos Monte Carlo

Uno de los métodos estadísticos más comúnmente aplicado en disciplinas de la ingeniería y de la ciencia en general, que se utilizan en muchos escenarios, desde la modelización financiera (nuestro caso) hasta los problemas de la física teórica. Es una clase de métodos numéricos que se basa en el muestreo aleatorio y distribuciones de probabilidad para explorar el problema dado.

Entre los numerosos beneficios del uso de los métodos de Monte Carlo para la simulación es que permite modelar sistemas que son demasiado complejos para una solución analítica. Como lo es la estimación del capital mínimo para empezar a operar con el sistema (o portfolio de los mismos) y que una de sus variables principales será la estimación del máximo *drawdown*.

Según este tipo de procedimientos, existen estudios sobre estadística/probabilidad que tratan de demostrar que no es aconsejable aplicar los resultados optimizados IS (la mejor combinación de parámetros analizada en el postprocesado) en la muestra que hemos llamado OOS porque siguen siendo datos pasados aunque se pretenda simular un supuesto ‘futuro’. En cambio otros estudios sostienen que no sólo se debe probar con una muestra OOS sino que además fortalece acompañar con una simulación Monte Carlo para estimar el riesgo con el muestreo aleatorio. Esto será lo que vamos a realizar durante el desarrollo del trabajo, vamos a realizar un estudio mixto para que un procedimiento (Monte Carlo) refuerce al anterior (OOS) para una determinada variable (Máximo *drawdown*).

* $r_{xy}^2 = \frac{s_{xy}^2}{s_x^2 * s_y^2}$ tomando covarianza y varianzas muestrales. Comprendido entre 0 y 1, preferiblemente cercano a 1.

Entre los posibles métodos Monte Carlo, vamos a aplicar el procedimiento explicado más abajo para que en base a los resultados de la simulación IS, seguidamente después del postprocesado de los datos para obtener la mejor combinación de parámetros, nos ofrezca así los resultados del *backtest* OOS. Con la *equity* completa de los datos IS y OOS, realizaremos el estudio Monte Carlo y así obtener el máximo *drawdown*. Cabe destacar que el dato máximo *drawdown* es la base para calcular el capital mínimo inicial antes de comenzar la operativa en tiempo real.

1. Primero se cogen las operaciones realizadas (*equity*) de la parte de la muestra IS y OOS.
2. Considero que el resultado del postprocesado (número de *trades*, para así representar la *equity*) con la muestra escogida OOS es representativa (suficientes datos) por el método estudiado.
3. Aplico permutaciones sin repetición.
4. Realizo una simulación con varias iteraciones para obtener un conjunto de curvas suficientemente representativa. Obteniendo así los máximos *drawdowns*.
5. Con un diagrama de barras o histograma se representan los máximos *drawdowns* obtenidos para estimar la posible distribución de probabilidad (generalmente Normal en estos casos). Es decir, la forma de la representación debe ser aproximadamente a la de la función densidad de probabilidad (fdp) que modela los datos.
6. Fijo los niveles de confianza $(1 - \alpha)$. Generalmente del 95%. Siendo α el nivel de significación.
7. Finalmente se obtiene el resultado, máximo *drawdown*.

1.4.4. Optimización

La optimización es un método cotidiano recurrido por la ciencia para obtener mejores rendimientos sobre el problema a solucionar. Pero todo investigador de un problema que desea ajustarlo al máximo posible (IS) para que a futuro (OOS) se aprecien rendimientos realmente buenos no deberá descuidar la teoría estadística/probabilística para realizar un correcto postprocesado de los datos, porque si no caerá en la sobreoptimización y muy posiblemente (me atrevo a decir en el 99% de los casos) el algoritmo sobreajustado (*overfitting*) en el pasado tendrá una evolución en tiempo real con resultados desastrosos. En el tema que nos ocupa hablamos de un tema muy serio en todo momento, dinero, que no se debe malgastar sino rentabilizar. No comparto la postura de descuidar estos estudios, por muy complicados que sean y por el inconveniente del capital inicial (clave para comenzar la operativa en tiempo real), porque dedicarme entonces a ahorrar dinero, gastarlo en un capricho, volver a ahorrar dinero y gastarlo nuevamente en el siguiente capricho no lo veo rentable.

Hasta ahora en todo desarrollo siempre hemos citado en algún paso el término simulación y ésta se podrá realizar con la optimización de parámetros con la muestra de datos seleccionada (en nuestro caso el precio de cierre de un histórico financiero).

También hemos hablado de un procedimiento para ajustar mediante mínimos cuadrados y una vez más, optimizando unos parámetros (ajustando) a la ‘nube de puntos’ (precios de cierre), aunque dicho ajuste no es el problema que vamos a estudiar de la optimización porque su objetivo no es predictivo sino de preprocesado de datos (preparar la muestra IS) para estimar ante posibles *gaps* y/o desigualdades entre proveedores de datos.

Un Sistema Algorítmico de *Trading* viene determinado por unos parámetros y éstos dependiendo del valor que tomen tendremos, en una simulación *backtest* o en una operativa real, unos resultados de beneficio/pérdidas (P&L), entre otros resultados. Lo normal en estos casos es optimizar los parámetros del sistema con la muestra que comentábamos IS, priorizando el orden de los resultados según un ratio (medida de rendimiento). Podríamos suponer que al concluir la optimización se debería escoger la combinación de parámetros más prioritaria pero vamos a ver que los resultados OOS no resultan ser los mejores (tomando en cuenta el riesgo y no sólo valor del ratio y beneficio). Por otro lado, volvemos con los detalles computacionales en el cálculo de los resultados de la optimización. El ordenador tiene sus limitaciones y podemos alargar en un tiempo bastante considerable la optimización debido a que el número de combinaciones posibles que se pueden dar en los parámetros y ésto unido a los cálculos de cada uno en una muestra IS de datos bastante larga si tratamos por ejemplo con tres años con un *timeframe* de minutos. La optimización puede tomar dos posibles vías de ejecución, según el algoritmo de optimización, una secuencial (probar todas las combinaciones, ‘fuerza bruta’) y otra por algoritmos genéticos (el objetivo es obtener lo que le indicamos al algoritmo antes de probar todas las combinaciones).

El algoritmo genético es uno de los procedimientos de optimización más empleado por la ciencia y que actualmente sigue siendo motivo de estudio (Proyectos Fin de Carrera y Tesis Doctorales). El procedimiento base de una generación sería el siguiente (siguiendo además un ejemplo):

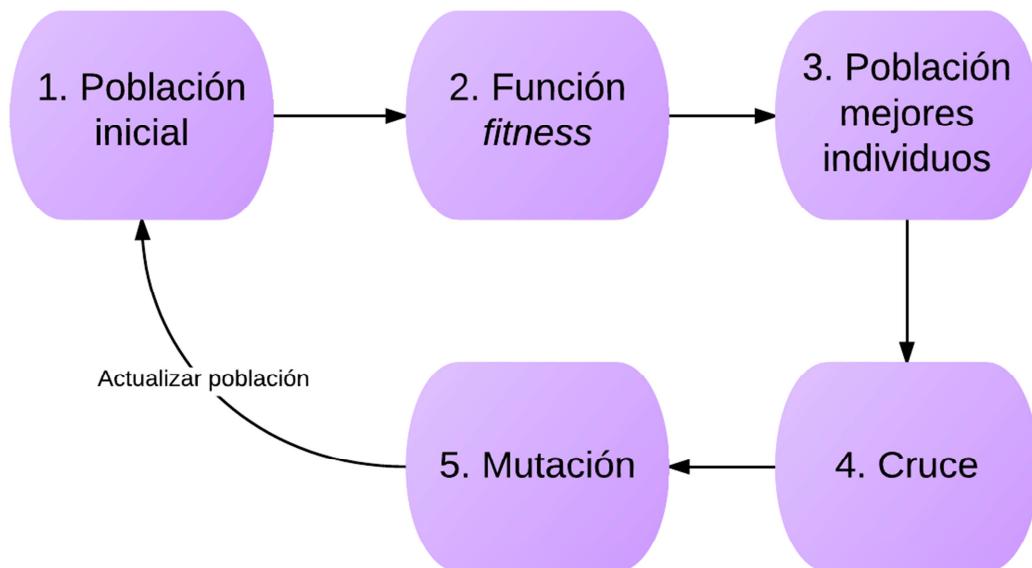


Figura 2: procedimiento básico de un algoritmo genético

1. Generar la población inicial

Según los experimentos de De Jong y Spears, la elección del tamaño de población es vinculante al resultado buscado. Lo típico se encuentra entre 10 y 100 individuos.

- Si el problema a resolver son operaciones sencillas -> población baja
- Si el problema es complejo -> población alta

El lector seguro que se estará preguntando qué es ‘baja’ y ‘alta’ población. Pues bien, como decíamos antes, este algoritmo todavía le queda estudio por resolver.

Ejemplo:

Tenemos tres variables a optimizar:

$$V_1 \in (1.00, 100.00]$$

$$V_2 \in [-2.50, 4.50]$$

$$V_3 \in [-1.00, 4.25]$$

Construimos la ecuación de la recta ($y = mx + b$), tenemos dos puntos, para normalizar las variables y éstas estarán comprendidas entre 0 y 1:

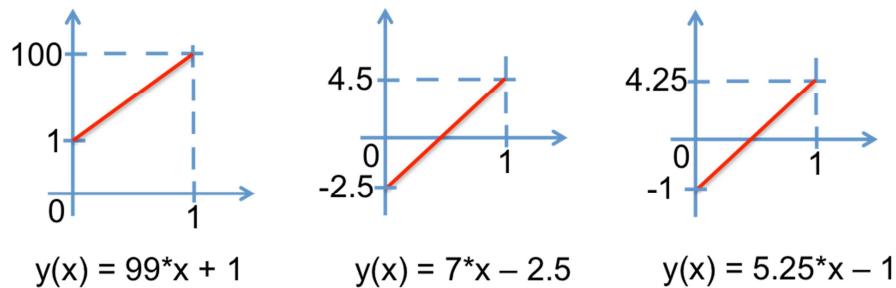


Figura 3: normalización de variables, algoritmo genético

Finalmente generamos la población de 5 individuos (n), para el ejemplo, de manera aleatoria. Será una matriz de 5 filas y 3 columnas ($V_1 \ V_2 \ V_3$):

$$\begin{bmatrix} 0.3 & 0.6 & 1 \\ 0.54 & 0.8 & 0.1 \\ 0.21 & 0.61 & 0.3 \\ 0.1 & 0.19 & 0.4 \\ 0.34 & 0.58 & 0 \end{bmatrix}$$

2. Implementamos la función *fitness* (F) para maximizar o minimizar durante la optimización. Continuamos con el ejemplo:

$$\text{Deseamos maximizar la función, } F = \frac{v_2 + v_3}{v_1}$$

Individuo tipo 1 [0.3 0.6 1] ----- $F_1 = 2.33$
 Individuo tipo 2 [0.54 0.8 0.1] ----- $F_2 = 1.67$
 Individuo tipo 3 [0.21 0.61 0.3] ---- $F_3 = 4.33$
 Individuo tipo 4 [0.1 0.19 0.4] ----- $F_4 = 5.9$
 Individuo tipo 5 [0.34 0.58 0] ---- $F_5 = 1.71$

3. Generamos, según su distribución de probabilidad, la población con los 5 mejores individuos otra vez de manera aleatoria. Continuamos con el ejemplo:

$$t = \sum_{i=1}^n F_i = 15.94$$

Individuo tipo 1 [0.3 0.6 1] ----- $F_1 = 2.33$ ----- $P_1 = 2.33 / 15.94 = 0.15$
 Individuo tipo 2 [0.54 0.8 0.1] ----- $F_2 = 1.67$ ----- $P_2 = 1.67 / 15.94 = 0.1$
 Individuo tipo 3 [0.21 0.61 0.3] ---- $F_3 = 4.33$ ----- $P_3 = 4.33 / 15.94 = 0.27$
 Individuo tipo 4 [0.1 0.19 0.4] ----- $F_4 = 5.9$ ----- $P_4 = 5.9 / 15.94 = 0.37$
 Individuo tipo 5 [0.34 0.58 0] ----- $F_5 = 1.71$ ----- $P_5 = 1.71 / 15.94 = 0.11$

Si representamos los P_i ($i = 1,2,3,4,5$) en un gráfico de tipo circular o tipo ‘tarta’ quedaría así:

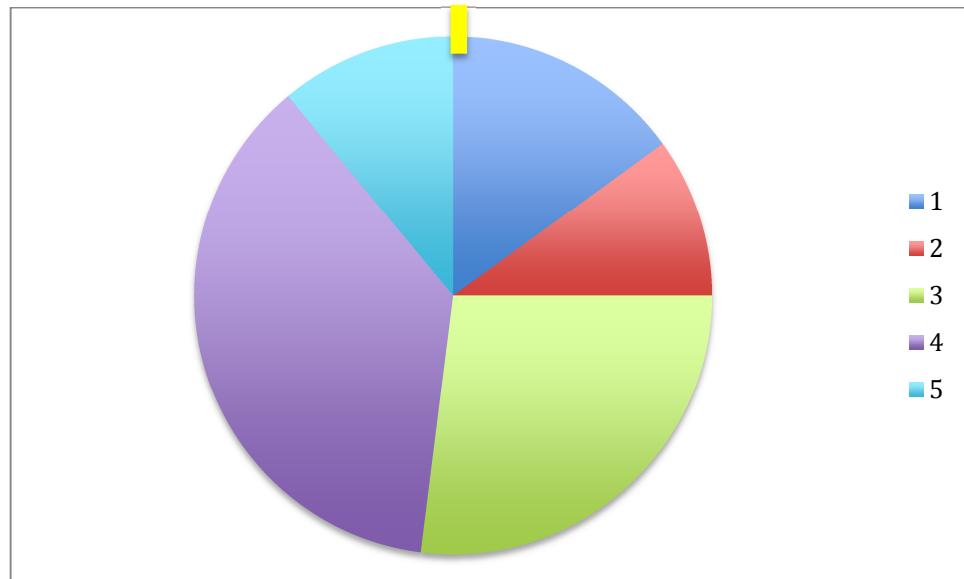


Figura 4: distribución de probabilidades según el valor de la función fitness

Para generar la población con los 5 mejores individuos, debemos hacerlo de manera aleatoria. Por tanto, supongamos que se trata de una ruleta (marca amarilla situada en las doce horas) y tiramos de ella 5 veces:

$$\begin{bmatrix} 0.1 & 0.19 & 0.4 \\ 0.1 & 0.19 & 0.4 \\ 0.21 & 0.61 & 0.3 \\ 0.3 & 0.6 & 1 \\ 0.1 & 0.19 & 0.4 \end{bmatrix}$$

Es lógico que la nueva población esté formada con más individuos del tipo 4, más probable, al ser la combinación de parámetros con una función fitness mayor.

4. Desarrollo de los cruces. Con los 5 mejores individuos establecemos los cruces entre los mismos según una probabilidad de cruce, establecemos las parejas de cruce y también seleccionando el punto de cruce. Continuamos con el ejemplo:

Suponemos que el punto de cruce lo realizamos entre la primera y la segunda variable. También suponemos que el cruce se realiza en parejas, recorriendo la matriz por filas una por una. Por lo general se toma una probabilidad de cruce mayor que la probabilidad de mutación, [25 - 95] %. Suponemos que en nuestro ejemplo se cruza el tercero con el cuarto individuo.

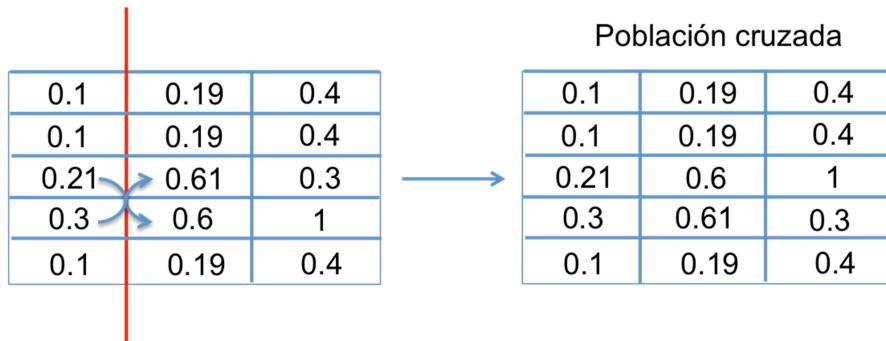


Figura 5: población cruzada

5. Mutación. Antes de terminar la generación se aplica un posible cambio del parámetro según una probabilidad de mutación dada (recorremos los elementos desde la primera fila hasta la última). El parámetro que debido a la probabilidad fijada sea candidato para mutar, aleatoriamente se le asigna un nuevo valor entre 0 y 1. Generalmente se toma como probabilidad el inverso del número de individuos de la población inicial y si encima queremos ser más precisos se puede unir un nuevo parámetro (ε) a optimizar para conocer cuál debería ser la probabilidad de mutación.

$$P_{mutación} = \frac{1}{n * \varepsilon}$$

Continuamos con nuestro ejemplo y suponemos que han mutado el parámetro 5 y el 12.

Población mutada		
0.1	0.19	0.4
0.1	0.19	0.4
0.21	0.6	1
0.3	0.61	0.3
0.1	0.19	0.4

Figura 6: población mutada

Con la población mutada ya hemos concluido la primera generación. Actualizamos la población inicial con la población mutada y repetimos el mismo proceso para la siguiente generación.

Por un lado tenemos los parámetros básicos para ejecutar el algoritmo:

- Población inicial.
- Función *fitness*.
- Probabilidad de cruce.
- Probabilidad de mutación.

Por otro lado tenemos como principales parámetros personalizados, a elección del programador, para parar o reiniciar el algoritmo:

- Número de generaciones antes de parar la ejecución.
- Tiempo transcurrido antes de parar la ejecución.
- Valor objetivo de la función *fitness* antes de parar la ejecución.
- Si el valor de la función *fitness* no mejora en un determinado número de generaciones, entonces parar la ejecución o generar nueva población para continuar ejecución.

Existen muchos derivados del procedimiento base explicado y que por supuesto vendrán marcados con sus ventajas e inconvenientes. En el presente trabajo no contemplamos como estudio el mejor algoritmo genético a utilizar, por lo explicado antes (podría abarcar un nuevo Proyecto Fin de Carrera o Tesis Doctoral), sino que nos centraremos en el estudio que nos ocupa de la sobreoptimización y lo veremos por medio de la fuerza bruta con muestras de combinación de parámetros aceptables para no alargarnos en el tiempo con las optimizaciones.

Antes de concluir, remarcar que difícilmente podremos hablar en términos poblacionales, en el punto 1.4.1. citamos ‘muestral’ para el cálculo de medidas durante el desarrollo del trabajo. Entre las posibles combinaciones de parámetros que pueda tener un sistema y luego la serie de datos histórica del producto financiero a analizar, que es bastante amplia si lo tratamos de manera intradiaria (ni que decir tiene lo que sería en HFT), no resulta ser el camino eficiente e inteligente probar toda la población con los ordenadores de uso particular. Sin olvidar que estamos estudiando un histórico que ya ocurrió (pasado) para estimar el posible comportamiento que vendrá (futuro), por ende podríamos tener *overfitting* y ésto no lo queremos, por supuesto.

1.5. Formulación del problema

Como suele pasar con cualquier modelo de comportamiento optimizado a pasado, el comportamiento a futuro se puede medir con estimaciones estadísticas/probabilísticas y éstas nos darán un porcentaje de casos ganadores y otro de casos perdedores. La optimización puede llevarnos al problema, si no se hace correctamente, del llamado *overfitting* (sobreoptimización, sobreajuste) cuando el Sistema Algorítmico de *Trading* se ‘pone a punto’ en la etapa de análisis mediante unos patrones definidos para ejecutarlo en el mercado. Todos los *traders* deseamos, como no, tener el mayor beneficio posible cuantificando el riesgo a asumir, también existe el problema de saber seleccionar la mejor función de *fitness* (función objetivo o medida de rendimiento generalmente cubierto por un ratio) para ordenar los resultados de optimización. Otro punto problemático es la obtención de los datos, las muestras para optimizar (IS) y las muestras para probar (OOS). Resulta que los proveedores de datos utilizan, como es lógico, servidores para el *streaming* de datos con el mercado electrónico y los clientes. Las tarifas además son variables según el tipo de datos que precisa el cliente. No todos los proveedores van a soportar las mismas infraestructuras de resguardo (servidores espejo y/o configuración de diversos RAID de almacenamiento) ante posibles caídas de conexión, caídas de tensión eléctrica o por sobrecalentamiento de los servidores, entre otros. Una empresa que opera con varios fondos de inversión, puede tener mayor riesgo si no filtra los históricos para crear una base de datos fiable a su juicio. Es más, existen compañías como Tick Data Inc. que se dedican a la recopilación de datos bursátiles para filtrarlos y finalmente construir, según ellos, el histórico bursátil 100% fiable. Este tipo de compañías no resultan ser nada baratas y pueden llegar a pedir cientos de dólares por el histórico de cada producto financiero.

Después de ver los posibles problemas de optimización, rendimiento y fuente de datos pasamos a ver otro problema crucial desde el punto de vista tecnológico y de cartera (para tener más potencia de cálculo -> más dinero que desembolsar). Para realizar este tipo de estudios, el *trader* necesita analizar grandes volúmenes de datos que computacionalmente hablando si no se dispone de conocimientos en programación en paralelo y de hardware, fácilmente se puede pasar el *trader* más de un mes optimizando (un ejemplo más de la función que realizaría un *quant*), años incluso, y el tiempo es oro. Este problema bien lo conocen las personas que se dedican a la criptografía (arte y ciencia de mantener un mensaje seguro) y/o al criptoanálisis (arte y ciencia de desvelar texto cifrado).

Con el objetivo de crear algoritmos de codificación más complejos para que las contraseñas no sean “rotas” con fuerza bruta (probar todas las combinaciones posibles). Como curiosidad, actualmente en codificación simétrica, para ‘romper’ una contraseña de 128 bits ($2^{128} = 3.4 \cdot 10^{38}$ número de claves distintas) con un millón de ordenadores se necesitarían $5.4 \cdot 10^{18}$ años (mientras no nos refiramos a los estudios actuales con ordenadores cuánticos que probablemente estén estudiando los de la NSA, si el ordenador trata con algo más que bits entonces ya es otra cosa). Eso sí, la contraseña se debe seleccionar correctamente (combinación de letras, números y símbolos) porque no vale el truco del tipo ‘\$eÇreT0’. Los criptoanalistas como mínimo te pasarán la prueba de diccionarios y los que se dedican a esto tienen el mismo o más ingenio que los que no nos dedicamos al criptoanálisis para crear combinaciones con reglas de construcción de palabras leídas con símbolos y números como lo haría una persona corriente para su fácil memorización.

Volviendo a los problemas que nos ocupa en el trabajo, remarcar que el estudio a realizar en su conjunto es complejo en la conclusión de resultados, realizaremos varias simulaciones para demostrar conclusiones y para que ésta sea posible tendremos que suponer algunos parámetros (parámetros que se citarán como tal para su fácil identificación). Para explicar la razón empírica de cada parámetro que supondremos podría conllevar perfectamente a la realización de un nuevo PFC o incluso Tesis Doctoral, en el campo de la optimización con sus resultados estadísticos/probabilísticos y sus tiempos de ejecución óptimos para cualquier campo de aplicación (robótica y *trading* entre otros) queda todavía mucho por aprender. Debemos tener en cuenta que las simulaciones que realizaremos no resultarán ser perfectas, es decir, la ejecución del sistema a mercado en tiempo real con una cuenta real, como mínimo estará sujeto según las operaciones que realice a las colas en la posición donde desea entrar comprado o vendido (debe existir demanda para vender y debe existir oferta para comprar. No está sólo en el mercado, hay más *traders* que desean entrar antes que usted). Generalmente los datos de tarifa básica proporcionados por los proveedores de datos, más adelante lo veremos, no ofrecen el dato de número de operaciones por posición (*tick*) y por tanto ya podemos confirmar que la prueba 100% fiable OOS sería en tiempo real con una cuenta real. Ahora bien, el precio que hay que pagar es empezar la operativa con la apertura de dicha cuenta con dinero real sin haber estado seguro al 100% (por ello hablamos en términos estadísticos/probabilísticos) de haber completado con éxito el estudio del análisis.

1.6 Objetivos

En el punto anterior hemos podido comprobar los principales problemas a los que se debe enfrentar una persona que se inicia en el campo del *trading* cuando se trata de analizar el pasado para estimar un futuro. A continuación vamos a mencionar, por el mismo orden expresado en el punto anterior, cuáles van a ser los procedimientos a demostrar para solventar los posibles problemas. Más adelante explicaremos el cómo de los procedimientos (metodología).

Objetivo 1: Demostrar que, independientemente de la función *fitness* empleada, se produce sobreoptimización si optimizamos el Sistema Algorítmico de *Trading* con un elevado número de combinaciones (posibles valores de los parámetros). Tomaré un rango de parámetros fijo, en este caso aquel que contenga el mayor número de combinaciones (10404) entre los tres diferentes rangos explicados en el siguiente objetivo. A continuación muestrearé de dos en dos teniendo entonces 5202 combinaciones y finalmente volveré a muestrear de dos en dos respecto a la anterior matriz o de cuatro en cuatro respecto a la primera matriz (10404 combinaciones) para tener así 2601 combinaciones. Como demostración gráfica, observaremos la *equity* IS y OOS para comprobar que los rendimientos empeoran fuera de la optimización (OOS) cuando mejoran mucho más dentro de la optimización (IS). Como demostración cuantitativa, obtendremos por experimento una probabilidad de sobreoptimización (PBO) para comprobar que a medida que aumenta ésta, los rendimientos fuera de la optimización (OOS) disminuyen.

Objetivo 2: Demostrar que podemos disminuir el riesgo dependiendo de la mejor combinación de parámetros seleccionada y que en base a la función *fitness* a optimizar, obtendremos más o menos rendimiento en función del riesgo mencionado. Para elegir la mejor combinación de parámetros voy a utilizar un procedimiento básico pero con fundamento. Realizaremos tres experimentos distintos (según el número de combinaciones) mediante optimización de ejecución secuencial (paralelizado para utilizar seis núcleos, cores, del ordenador), con un histórico financiero preprocesado, dos posibles funciones *fitness* (cada experimento se realiza con una función *fitness* o medida de rendimiento, por tanto llegados a este punto tendremos un total de seis experimentos más el experimento del objetivo 1, un total de siete), y un Sistema Algorítmico de *Trading* de uso frecuente. Representaremos una distribución de frecuencias con los intervalos que fijaremos más adelante según el ratio utilizado (función *fitness*) en la optimización, así observaremos medidas de tendencia central y de dispersión. Por otro lado utilizaremos el ajuste por mínimos cuadrados para representar la regresión lineal de la *equity* y que gracias al coeficiente de determinación calculado nos ayudará a seleccionar la evolución de la *equity* lo más lineal posible. Cuanto más lineal es la *equity* se suele decir que el sistema es más robusto. En toda estimación estadística siempre se comenta si la muestra es lo suficientemente significativa como para dar los resultados por válidos, como vamos a tratar con un sistema intradiario (*timeframe* en minutos) y de una muestra IS de varios años, aplicaremos para ello como filtro tres casos (50, 150 y 300 *trades* como mínimo). Por último realizaremos una simulación Monte Carlo con la *equity* seleccionada como mejor combinación de parámetros para obtener el máximo *drawdown*. Como demostración gráfica, observaremos la *equity* IS y OOS para comprobar los rendimientos. Como demostración cuantitativa, obtendremos por experimento (llegados a este punto, diecinueve experimentos. $3^2 \cdot 3 + 1$) el ratio Calmar para comprobar que a medida que aumenta éste, el riesgo disminuye y que el mayor ratio Calmar nos resolverá la elección de la medida de rendimiento (dos posibles, función *fitness*, como ya hemos mencionado) y mínimo número de *trades* para utilizar como filtro.

Como el presente trabajo no trata de demostrar en particular cuál es el mejor estudio para seleccionar la mejor combinación de parámetros (sólo un Proyecto Final de Carrera o Tesis Doctoral podría abarcar este único tema), recomiendo añadir posiblemente al proceso más pasos para estar más confiados de que se ha seguido un riguroso estudio y, como no, mayor número de experimentos si se dispone de tiempo para ello. Como es necesario presentar resultados al concluir los experimentos, en vez de suponer la elección de la mejor combinación de parámetros, se procede con un procedimiento con fundamento como ya he dicho.

Objetivo 2 bis: De lo comentado en el anterior objetivo, podremos obtener también otra conclusión. El máximo *drawdown* de la muestra del histórico financiero (IS y OOS), a razón del postprocesado de los datos para elegir la mejor combinación de parámetros, será menor que el máximo *drawdown* calculado en el experimento Monte Carlo. Demostraremos que gracias al procedimiento Monte Carlo, adquirimos más seguridad/tranquilidad, por el estudio realizado a pasado, para estimar el dinero necesario de la cartera inicial para empezar la operativa real. Como lo importante es reducir el riesgo, el experimento Monte Carlo lo calculamos sobre la muestra total, considerando IS y OOS porque aunque lo normal sea que el máximo *drawdown* OOS sea mayor que en el período IS, siempre pueden ocurrir las excepciones. Lo veremos también en los resultados.

Objetivo 3: Demostrar rigor y mayor seguridad/tranquilidad de los resultados obtenidos cuando se realiza un preprocesado de datos para filtrar todas las posibles irregularidades del histórico financiero y finalmente aplicar un procedimiento para construir el histórico continuo aplicando un correcto *rollover* (con el ajuste de vencimientos, más conocido como *split*) y así aplicarlo en los experimentos a realizar. Sin olvidar que gracias al procedimiento de mínimos cuadrados (ajuste del continuo construido con una curva polinómica) podremos separar los resultados según tendencia alcista | bajista | lateral, ésto nos permitirá conocer en qué área/s se comporta mejor el Sistema Algorítmico de *Trading*. Aparte de comparar los históricos de dos proveedores y de comprobar que se ajustan a las características del producto financiero a estudiar (en nuestro caso el futuro del E-mini S&P 500), con el procedimiento que describiremos más adelante podremos estar seguros en la medida de lo posible del histórico continuo que utilicemos para optimizar los experimentos necesarios. Nombraremos varios métodos de construcción del citado continuo, razonaremos qué método utilizaremos y finalmente aplicaremos dicho continuo ajustado en los dieciocho experimentos nombrados en el objetivo 2. Si no aplicásemos un razonamiento de ajuste del histórico continuo, los resultados de los experimentos vendrían sesgados, contando con que el rendimiento podría decrementarse o incrementarse significativamente y para demostrarlo añadiremos los experimentos veinte y veintiuno. Tanto el experimento veinte como el veintiuno se elaboran con el mismo proceso según el considerado como mejor en los anteriores dieciocho experimentos del objetivo 2, sólo se distinguen porque en uno (experimento veinte) se empleará el proveedor1 y en el otro (experimento veintiuno) el proveedor2. Otra posible conclusión que obtendremos será que los resultados serán distintos tanto por proveedor como en lo obtenido en el mejor de los dieciocho experimentos del objetivo 2 y demostraremos lo explicado en el punto anterior referente a la fiabilidad de los datos.

Como ya hemos anunciado, este experimento consistirá en: Después de escoger el mejor experimento (beneficio/riesgo) de los dieciocho posibles (objetivo 2), éste lo volveremos a realizar pero esta vez sin un histórico continuo elaborado, sino uniendo simplemente vencimientos (rollover con fecha fija, explicado en el punto 2.5) y sin aplicar tampoco un ajuste (*split*), tanto en el experimento veinte (proveedor1) como en el experimento veintiuno (proveedor2). Puede parecer estúpido, después de lo poco explicado sobre históricos de datos y vencimientos de futuros (hablaremos en detalle más adelante), no aplicar un *rollover* y *split* razonado del histórico pero ¿se ha preguntado alguna vez cómo construye el histórico continuo la plataforma de trading que utiliza?

Objetivo 4: Los temas relacionados con la computación (tanto hardware como software) son de especial importancia. Como mínimo la ejecución de todos los experimentos se realizarán, como ya se ha anunciado, en paralelo en función de las características hardware del ordenador (más adelante lo detallaremos). Veremos que a pesar de no utilizar una muestra de combinación de parámetros tan grande si se apreciarán diferencias computacionales si lo escalásemos a un problema con mayor número de combinaciones. Como trabajo extra para lidiar con los rendimientos computacionales, realizaré varias pruebas en el experimento veintidós. Ejecutando así la mayor combinación de parámetros de los anteriores dieciocho experimentos (objetivo 2) de forma no en paralelo y también sin reservar memoria como aconsejábamos para comprobar diferencias computacionales.

En resumidas cuentas, podemos englobar todos los objetivos mencionados en:

- Conseguir una base de datos más fiable y segura para procesar los datos en simulación (experimentos). También comprobamos en qué área/s se comporta mejor el Sistema Algorítmico de *Trading* (tendencia alcista | bajista | lateral). Además veremos las comparaciones de resultados en rendimiento según el proveedor de datos seleccionado.
- Mediante el cálculo de la probabilidad de sobreoptimización (PBO), unido con el procedimiento básico descrito para calcular la mejor combinación de parámetros, conseguir un rendimiento beneficio/riesgo más fiable y seguro.
- Conseguir mejorar el rendimiento computacional con la ejecución en paralelo.

2. Estado del arte

Hasta ahora hemos visto las ventajas e inconvenientes a las que se deberá enfrentar una persona que comienza en el área del *Algo Trading* cuantitativo, parte de la actividad de los mercados financieros. Hemos mencionado sistemas, optimización y medidas de rendimiento, entre otros, al citar los problemas y objetivos a resolver en el presente trabajo. Pues bien, a continuación es necesario detallar los agentes principales que intervienen en medio de las herramientas más habituales ya mencionadas.

2.1. Intermediarios y mercados financieros

Para poder realizar el estudio del presente trabajo, ya sea un preprocesado de datos, una optimización o un postprocesado, es preciso obtener primero por medio de alguna fuente los datos a procesar. Las fuentes de datos se obtienen de intermediarios financieros, generalmente *brokers* y proveedores exclusivamente de datos entre otros. Si en cambio dispone de una amplia cartera, el mismo regulador del mercado puede ofrecerle también datos (normalmente a través de la API de Bloomberg) y éstos lógicamente vendrán dados con mayor nivel de detalle que los datos típicos que utilizaremos y que éstos están disponibles además en varias páginas web (luego ya que sean fiables es otra cosa). La fiabilidad de los datos, sean del nivel de detalle que sean, cuesta dinero y más todavía si no conoces técnicas de programación para filtrarlos.

Los intermediarios financieros más típicos, con los que trabajaremos, son *brokers* (proveedor1) y empresas de desarrollo de plataformas de *trading* (proveedor2) que disponen de contratos con terceros (como Bloomberg) o con el mismo regulador del mercado (BME y CME Group entre otros) para ofrecer tanto datos históricos (con los que trabajaremos) como datos en tiempo real de productos financieros. Aprovecho para agradecer, una vez más, el poder recibir el material necesario por parte del proveedor1 y proveedor2 para elaborar el presente Proyecto Fin de Carrera. Al tratarse de empresas relevantes en el área de los mercados financieros y como realizaremos varios test de comparación respecto a sus datos, por respeto a las mismas no vamos a mencionar sus nombres.

Es muy importante recordar que en el punto 1.1 comentábamos "...la operativa en productos financieros se ha automatizado por completo...". Es decir, gracias a la automatización de los mercados financieros podemos disponer de históricos de datos con mayor o menor detalle (según lo dispuesto a pagar) para realizar estudios cuantitativos y éstos aplicarlos a estrategias algorítmicas para que la operativa sea 100% automática. En nuestro caso, utilizaremos los datos típicos, FECHA | HORA | APERTURA | MÁXIMO | CIERRE | MÍNIMO | VOLUMEN. Como he comentado, según el dinero, se podría obtener datos más precisos sacados del libro de órdenes. En dicho libro se reflejan los movimientos de oferta (bid), demanda (ask), precio del mercado (price), posiciones compradoras, posiciones vendedoras, colas (orden de llegada de las posiciones, operaciones, que se desean abrir) y seguro que de algún dato más que un profesional que se dedica al *High Frequency Trading* conoce mejor que yo.

La automatización de los mercados financieros y por ende la llegada de los mercados electrónicos permiten a cualquier persona con conexión a internet poder conectarse al instante. Pero ¿qué se necesita para operar en los mercados financieros electrónicos? El caso más típico sería abrir una cuenta de dinero con un *broker* (por lo general exigen dinero mínimo para su apertura). Lo normal sería que el *broker* disponga de dos conexiones al mercado, una a través de una plataforma de trading (aplicación local) que deberá descargar el cliente para su ejecución (licencia login facilitado por el *broker*) y por si falla algo en la plataforma, el *broker* dispone de una aplicación web para operar en el mercado con la misma cuenta contratada. El *broker* por lo general cobrará una comisión por vuelta en la operación realizada (abrir o cerrar un trade) y por los datos en tiempo real de productos financieros. Los datos históricos suelen venir en el pack contratado hasta un límite si hablamos de un *timeframe* = *tick* (6 meses por lo general). Por otro lado, cada vez que entramos a mercado ‘comprados’ o ‘vendidos’ disponemos de un apalancamiento, el regulador del mercado nos exigirá depositar unas garantías (dinero mínimo) ante posibles pérdidas en la operación. El regulador toma como responsable al intermediario financiero y lo normal será que si la posición abierta supera en pérdidas el equivalente a las garantías depositadas, el intermediario financiero cerraría automáticamente la operación cubriendo así la pérdida con las garantías obtenidas de la cuenta del cliente.

¿Por qué se exigen unas garantías? Por el citado apalancamiento que presenta el mercado de futuros. Las garantías, como decíamos, establecen el dinero mínimo requerido para operar con un mercado de futuros. Además estas garantías serán variables.

Cuando abrimos un contrato, según el intermediario financiero, se toman en cuenta todas o algunas de las siguientes;

Intraday Initial = garantías necesarias en el momento de abrir, por contrato.

Intraday Maintenance = garantías necesarias para mantener una posición, por contrato.

Overnight Initial = garantías necesarias en el momento de cerrar el activo subyacente (S&P 500, Stock) pero la posición en el derivado futuro sigue abierta.

Overnight Maintenance = garantías necesarias en el momento de cerrar el derivado financiero (futuro E-mini S&P 500) para mantener posición abierta.

Como el riesgo de mantener una posición abierta aumenta a medida que se cierra el mercado subyacente y posteriormente el derivado, lo más probable es que las garantías exigibles aumenten.

Un ejemplo, el futuro E-mini S&P 500 tiene un *tick* de 0.25 y 1 punto = 50 \$ (Símbolo establecido en el CME Group = ES | Según la plataforma de *trading* utilizada, el símbolo podría variar)

Si entramos con 1 contrato en 1800 puntos, su valor nominal sería $1800 * 50 = 60000$ \$ pero en vez de depositar toda la cantidad te permiten un apalancamiento y éste puede variar según el *broker*, la temporada y la volatilidad entre otros.

Si el intermediario financiero cobrase hoy 2703.125 \$ (Intraday Initial) por abrir una posición (contrato), quiere decir que tienes un apalancamiento de $60000/2703.125 = 22.197$. Es decir, 22.197:1. Por cada dólar que depositas, te permiten operar con 22.197 dólares.

Actualmente el denominado ‘futuro’ se encuentra como uno de los productos financieros existentes más populares y será con el que trabajaremos. También disponemos del mercado FOREX (divisas), stock (acciones), opciones, ETFs y CFDs entre otros. Sin olvidar que cada producto financiero está formado por cientos de mercados disponibles para operar.

Elijo el producto futuro porque; 1. Llevo trabajando con él, entre otros, desde hace más de siete años (prohibido operar con un producto sin conocer su funcionamiento), desde que mi primer mentor (mi abuelo) me lo enseñó, 2. Se trata de un mercado regulado (la operativa entra a mercado y no hay posibles ‘libros B’ de intermediarios financieros), 3. La operativa está consolidada en el conjunto global y por tanto suelen ser bastante líquidos (el seleccionado para el trabajo lo es) 4. Se puede operar tanto al alza (comprado) como a la baja (vendido) y 5. Tiene un apalancamiento bastante alto si no eres un profesional 100% de la materia.

El origen de los futuros como comentábamos en el punto 1.1, surgió a razón de los productores de materias primas para asegurar la venta de la cosecha a ‘fin de temporada’ (técticamente hablando se conoce como ‘vencimiento’). Actualmente existen mercados de futuros sobre índices (como el futuro del IBEX 35), acciones (como el futuro de la acción del banco Santander), divisas (como el futuro del par EURO/DOLLAR) y materias primas, como no. Se dice que el producto financiero futuro es

*“un contrato o acuerdo que obliga a las partes contratantes a comprar o vender un número determinado de bienes o valores (activo subyacente) en una fecha futura y determinada, y con un precio establecido de antemano. Estos contratos se negocian en lo que se llama mercado a término o mercado de futuros.” Wikipedia (**contrato de futuros**).*

Es decir, el activo subyacente es la referencia del derivado (‘futuro’ en este caso).

Pero entonces, si un operador bursátil abre una cuenta con un *broker* y opera en el mercado futuro del petróleo, ¿qué pasaría si abre un contrato (comprado o vendido) en una posición, a un precio (*tick*) y al llegar el vencimiento estipulado por el regulador de mercado, el operador no cierra dicho contrato? Teóricamente, si compró por ejemplo dos contratos del mercado futuro del petróleo en un vencimiento dado y termina éste venciendo, el *broker* debería entregarle la proporción en barriles de petróleo que le corresponden al cliente. Habrá notado el lector un poco de ironía en el ejemplo, como es lógico el intermediario financiero (*broker* en este caso), ‘se cura en salud’ y cierra posiciones del vencimiento pasado (si no lo hizo el cliente) para volverlas a abrir en el vencimiento posterior. A esta práctica se la conoce como *rollover*.

En la misma situación se encuentran las divisas, en vez que el intermediario financiero te ofrezca por derecho los dólares que compraste, se realiza el *rollover* por lo mismo de antes.

Como hemos comentado, existen varias decenas, incluso cientos de mercados según el regulador de mercado. En este caso vamos a centrar el estudio en el mercado futuro del E-mini S&P 500. Fíjese que se trata de un mercado ‘mini’, es decir sus características de operativa son más bajas (comisiones, garantías, valor del punto en divisa) que el mercado padre (también se han creado los mercados ‘micro’). En este caso, el mercado citado está regulado por el CME Group y su ficha técnica completa se puede encontrar en el ANEXO I.

Los datos más importantes que utilizaremos para realizar los cálculos serán:

- 1 punto = 50 \$
- El tamaño del *tick* es 0.25. Es decir, la evolución del precio, si éste subiera, sería 1890.25 -> 1890.50 -> 1890.75
- El vencimiento es el tercer viernes del mes a las 15:30 horas (hora local, UTC+1 horario CET y UTC+2 horario CEST). Los meses de vencimiento son marzo | junio | septiembre | diciembre.
- Según las condiciones contratadas con el intermediario financiero (datos reales sin citar la empresa en cuestión):
 - Horario de negociación (el horario electrónico se extiende más aun pero en el estudio vamos a operar dentro del horario del activo subyacente): 8:30 – 15:15 (ETH). Es decir las 15:30 – 22:15 (hora local, UTC+1 horario CET y UTC+2 horario CEST)
 - Comisión por vuelta: 0.85 \$ ($1.7 \frac{\$}{trade}$)
 - Garantías iniciales (**Intraday Initial**): 2875 \$
 - Garantías de mantenimiento (**Intraday Maintenance**): 2300 \$
 - El resto de garantías no las vamos a considerar porque el Sistemas Algorítmico de Trading que veremos cierra posiciones antes de cerrar el mercado subyacente.

2.2. Indicadores Algorítmicos de uso frecuente

Uno de los principales desarrolladores de indicadores técnicos para utilizarlos en la operativa fue J. Welles Wilder Jr. Comenzó su vida laboral como ingeniero mecánico pero su principal pasión era la investigación de los mercados financieros y da la casualidad que el producto futuro le interesaba especialmente. Hay que considerar que los indicadores típicos, que ahora son conocidos e impartidos en cualquier curso de formación sobre trading (de pago o gratuito), fueron estudiados, podríamos decir, en otras condiciones de mercado. Es decir, antes de la llegada de los mercados electrónicos, abarcando todas las ventajas que ya hemos expuesto, fueron desarrollados los indicadores por Wilder.

Si ahora con las máquinas que podemos disponer y con los programas informáticos nos parecen pocos los datos nombrados (porque lo son, mínimo para la persona que se dedica en el *High Frequency Trading*) para realizar una simulación perfecta (tomando en cuenta todas las condiciones de mercado según el libro de órdenes), imagínese antes de todo esto cuando sólo se disponía de la apertura y cierre del día. Remarco que la evolución de los mercados financieros está siendo muy notable para abrir nuevos cauces de estudio y que la comunidad científica e ingenieros en general sean ahora los más demandados para analizar e identificar patrones. Los indicadores más conocidos que se le atribuyen a J. Welles Wilder Jr. son el RSI (Relative Strength Index) y el ATR (Average True Range).

Para el Sistema Algorítmico de *Trading* (estrategia para la operativa) básico, como herramienta a estudiar en el proceso de análisis, vamos a emplear dos medias móviles. El indicador de media móvil se puede basar en varios tipos, en este caso aplicaremos por un lado una media simple (aritmética de toda la vida) y por otro lado la media exponencial.

La nomenclatura más conocida para la media simple es SMA (Simple Moving Average). El indicador dispone de un parámetro, el período de datos (ventana) para calcular dicha media. Realizaremos el cálculo de la forma más típica, escogeremos el precio de cierre del histórico y optimizaremos cuántos datos de cierre se necesitan para que la estrategia funcione. Del mismo modo pasará con la media exponencial, EMA (Exponential Moving Average), tendremos también un único parámetro.

Aprovechando que he estudiado la licenciatura de Ingeniero de Telecomunicación podemos implementar las medias aplicando los conocimientos estudiados en filtros digitales.

SMA es lo mismo que hablar de un filtro FIR (*Respuesta al Impulso Finita*):

$$a(1)y[n] = x[n] + x[n-1] + \dots + x[n-N_{sma}-1]$$

$$a(1) = N_{sma}$$

N_{sma} es el período del filtro a determinar.

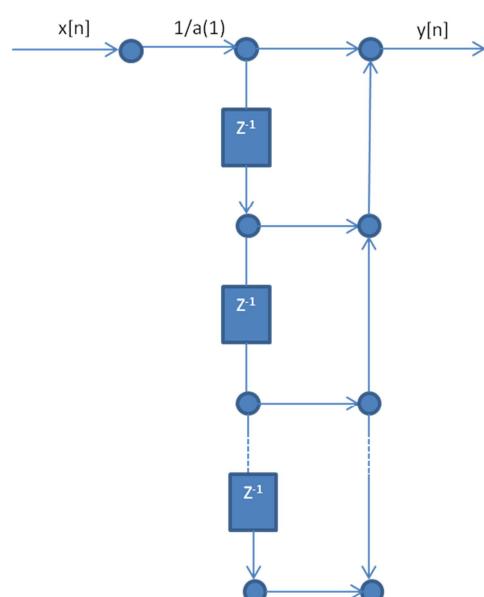


Figura 7: flujograma del filtro FIR-SMA en su forma canónica o tipo II

EMA es lo mismo que hablar de un filtro IIR (Respuesta al Impulso Infinita):

$$a(1)y[n] + a(2)y[n-1] = b(1)x[n]$$

$$a(1) = 1 \quad a(2) = \frac{N_{ema}-1}{N_{ema}+1} \quad b(1) = \frac{2}{N_{ema}+1}$$

N_{ema} es el período del filtro a determinar.

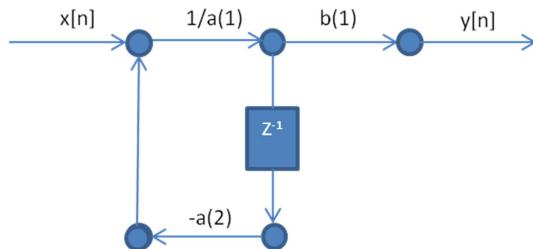


Figura 8: flujograma del filtro IIR-EMA en su forma canónica o tipo II

2.3. Sistema Algorítmico de *Trading* típico

Normalmente la ejecución de los sistemas los ejecuta un *trader* independiente en su PC personal. Para evitar el riesgo de apagones y demás problemas que se puedan presentar, existen empresas tecnológicas (como *SpeedyTradingServers*) con un despliegue de servidores y con todas las medidas de seguridad necesarias (backup, refrigeración y asistencia 24/7 entre otros) para la correcta monitorización del sistema. A este servicio se le conoce como *Virtual Private Servers* (rama de los famosos servicios *cloud*). Pueden ser varias máquinas virtuales compartiendo un servidor o un servicio de servidor dedicado, lógicamente éste último es más caro. Si se dispone de una cartera bastante grande (no es mi caso), varios cientos de miles de euros, podríamos ejecutar los Sistemas Algorítmicos de *Trading* en los servidores que dispone el mismo regulador del mercado o llevar los servidores personalizados al recinto que dispone el regulador. A esta modalidad se le llama *colocation* y es muy necesaria para aquellos fondos de inversión que estén operando en *High Frequency Trading*, una de las ventajas más cotizadas se basa en ser el más rápido para entrar en el mercado.

Volviendo a los avances estudiados por Wilder, a pesar de no disponer de datos intradiarios, pudo realizar estrategias consistentes en el tiempo (diario) para operar en el mercado. Ahora pretendemos, los que empezamos, utilizar indicadores ‘de la época’ para una estrategia intradiaria, en mi opinión el riesgo es muy alto y lo veremos con los resultados obtenidos. Lo normal es aplicar más procedimientos en la estrategia, ahora se barajan los sistemas que se adaptan con el aprendizaje (algoritmos *Machine Learning*).

Lo que si defiendo al 100%, dejando a un lado el tipo de estrategia a emplear, es el análisis de un preprocesado y postprocesado de datos (labores de un *quant*), básicamente lo que pretendemos demostrar con este trabajo aunque todavía me faltan por estudiar más pasos del desarrollo como comentaba en el punto 1.5 y como seguiré anunciando a lo largo de lo que queda de trabajo. Por cierto, de otra cosa estoy 100% seguro, la perfección no la puede conseguir la raza humana y que, sin hacer las típicas trampas que tanto se repetirán a lo largo de la historia (información privilegiada, chivatazos) no se puede tener probabilidad 1. Un buen analista concluirá en base a un análisis, la probabilidad de que el suceso ocurra. ¿Mañana subirá o bajará el futuro IBEX 35?

Para que podamos concluir con resultados sobre el proceso de análisis a estudiar, vamos a utilizar un sistema con una estrategia muy típica, el cruce de dos medias móviles. Cualquiera que empiece en el mundo del *trading*, ya sea un curso de pago o gratuito, es formado con la estrategia del cruce de medias pero hay que tener cuidado en aquel ‘profe’ que se gana la vida con el curso del cruce de medias pero no por el *magnific* procedimiento del cruce de medias. Este trabajo muestra un procedimiento de estudio que todo curso de pago debe enseñar, conceptos que se aprenden de entrada en una carrera universitaria de ingeniería y claro, las empresas necesitan una seguridad (control de riesgo) antes de invertir un millón de euros en un producto u otro y más aun un mileurista que por ejemplo reúne diez mil euros para comenzar a invertir.

Tenemos cuatro parámetros a optimizar:

- Período de la media simple (SMA), N_{sma}
- Período de la media exponencial (EMA), N_{ema}
- Hora comienzo de la operativa, H_c
- Hora fin de la operativa, H_f (siempre cerramos posibles posiciones abiertas antes de finalizar el día de negociación)

En el modelo las llamaremos: periSMA, periEMA, hComienzo, hFin

La estrategia es simple;

- Si EMA cruza por encima de SMA, el sistema cerrará posición si estaba vendido y comprará un contrato.
- Si EMA cruza por debajo de SMA, el sistema cerrará posición si estaba comprado y venderá un contrato.
- La operativa se realizará dentro del intervalo marcado por la ‘hora comienzo’ y ‘hora fin’.
- Las operaciones se ejecutan a mercado. La razón es debida a la poca resolución de los datos, lo explicaremos más adelante, por ello que si los resultados van a salir con cierto sesgo ya ni que decir tiene utilizar órdenes en stop o limitadas.

También existen otro tipo de operaciones a ejecutar. Las operaciones de tipo limitada se ejecutan con el cálculo en puntos o en porcentaje del precio según un posible umbral (parámetro a optimizar) antes de su ejecución. Las operaciones de tipo *stoploss* o *stop* se suele emplear para proteger la operación ya abierta como comprado o vendido con el parámetro a optimizar (umbral). Si después de comprar (como es lógico deseamos que suba el precio) un contrato y de lanzar un *stop* de venta, el precio empezase a bajar hasta sobrepasar el umbral donde se encuentra el *stop*, entonces se cierra la operación con dicho *stop*. Un derivado del *stop* es el llamado *trailing stop* que siempre acompaña, manteniendo la misma distancia, con la operación ya abierta (comprado o vendido).

Una imagen vale más que mil palabras, a continuación podemos observar que la línea azul es el indicador EMA y la línea roja el indicador SMA. Las líneas verticales de color rojo indican ventas y las líneas verticales de color verde indican compras.

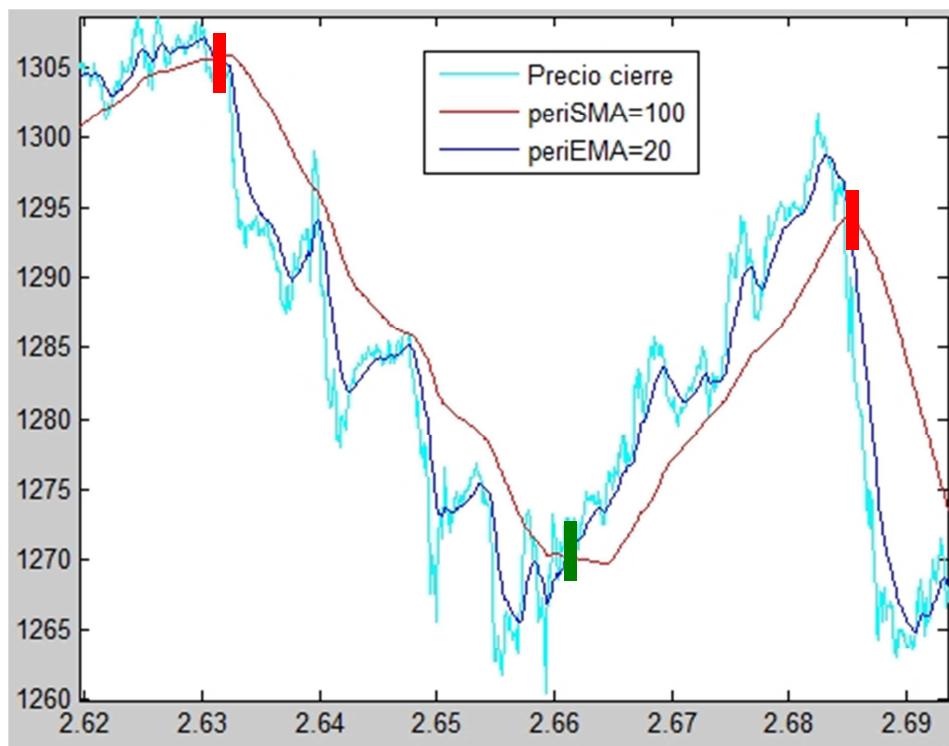


Figura 9: Ejemplo cruce de medias SMA y EMA

2.4. Medidas de rendimiento

En toda simulación, como veníamos anunciando, necesitamos conocer la función *fitness* que determinará si los resultados son favorables o no. La medida de rendimiento nos permite así ordenar los resultados y en particular para el procedimiento de estudio que realizo vendrá determinado por un ratio. Existe una gran variedad de medidas de rendimiento pero nos centraremos en las más importantes, aquellos ratios que más se utilizan en el mundo cuantitativo financiero.

Por otro lado consideraremos también medidas de rendimiento computacionales, muy importantes a considerar después de los argumentos expuestos en el punto 1.5.

La complejidad del asunto comienza en saber seleccionar el ratio idóneo para el estudio a simular. Muchas veces me pregunto, como otras personas, el por qué de las decisiones tomadas. Cuando he tenido la oportunidad de preguntar a investigadores que se dedican por cuenta propia al análisis de sistemas, o incluso formando parte de alguna empresa del sector financiero, suelen comentarme unos que utilizan un ratio concreto porque es el que mejor resultados le ha dado desde hace varios años o quizás porque directamente es el más utilizado por la comunidad. El problema es que existe escasez de base científica para confirmar que un ratio es mejor que otro para simular con tal sistema.

Pues bien, este es uno de los cometidos de este trabajo, como comentábamos en el objetivo 1 del punto 1.6, demostrar con un procedimiento razonado qué ratio escoger. El ratio más comentado por la comunidad para su uso es el llamado *Ratio Sharpe* pero existe otro ratio también muy defendido llamado *SQN*. Ambos ratios, lo veremos en sus fórmulas, son muy parecidos.

La fórmula general del Ratio Sharpe es $S_p = \frac{\mu - r_f}{\sigma}$

μ = media de los rendimientos en un período

σ = desviación típica de los rendimientos en un período

r_f = rendimiento libre de riesgo

Generalmente se toma como rendimiento libre de riesgo los bonos del tesoro pero no conozco a nadie, ni yo apuesto por ello, que me confirme que exista un producto financiero, sea del apalancamiento que sea, libre de riesgo. Por eso analizamos con procedimientos cuantitativos una estimación para reducir al máximo posible el riesgo, lo que desea escuchar un inversor es algo así como “No vas a perder más de un tanto por ciento del capital invertido en un período dado de operativa”. Por tanto volvemos a escribir la fórmula del Ratio Sharpe:

$$SR(\text{normalizado}) = \frac{\mu}{\sigma} * \text{anualizar}$$

Anualizar = $252 * 60 * 20 / \text{timeframe}$

Días de negociación == 252

Horas al día == 20

Minutos durante el día == $60 / \text{timeframe}$

A continuación veamos de qué variables está formado el Ratio SQN (System Quality Number);

$$SQN = \frac{\mu}{\sigma} * \sqrt{N_{cont}}$$

μ = media de los rendimientos por operación (*trade*)

σ = desviación típica de los rendimientos por operación (*trade*)

N_{cont} = Número de operaciones (*trades*) realizadas en total.

Como podemos observar, el *Ratio SQN* es prácticamente el mismo que el SR. Si acaso el SQN aporta más información respecto al número de *trades* realizados y el período de los resultados de dispersión varía. ¿En qué podría beneficiarnos realizar un cálculo con mayor número de operaciones? Como bien expresábamos en el objetivo 2 del punto 1.6, "...en toda estimación estadística siempre se comenta si la muestra es lo suficientemente significativa como para dar los resultados por válidos". Debido a que son bastante parecidos y la comunidad está dividida más o menos en un 50/50 utilizaremos estos dos ratios como medidas de rendimiento para compararlos.

En resumidas cuentas, enfrentaremos en los experimentos a realizar los ratios SQN y SR. Considerando además que a los dos les vamos a aplicar un filtro, marcado también en el objetivo 2 del punto 1.6, sobre número de contratos mínimos a realizar. Resulta que el ratio SQN nos está aportando como mejores resultados aquellos que realizan mayor número de *trades* pero para comprobar, aun así, que la muestra es significativa vamos a proceder filtrando con las tres siguientes pruebas: 50 | 150 | 300, número de *trades*. Si al finalizar el estudio resulta que los resultados beneficio/riesgo son mejores o muy parecidos con un SR y un filtro de 150 *trades* mínimos en contra de un SQN con un filtro de 150 contratos, entonces quizás el SQN no aporta más rendimiento de lo que sostienen sus defensores.

Antes de concluir con las medidas de rendimiento cuantitativo, exponemos como última medida de rendimiento a considerar en el trabajo el *Ratio Calmar*.

$$CR = \frac{B}{M_{dd}} \text{ (aplicado a 3 años)}$$

B = máximo retorno

M_{dd} = máximo *drawdown* (siempre hay)

El *Drawdown* es el retroceso de la curva de resultados (*equity*). A la curva de resultados también se la conoce como *Equity Line*. Generalmente se expresa en la divisa correspondiente aunque también se puede mostrar en porcentaje.

El CR es muy utilizado también en el sector financiero, por ejemplo lo utiliza para puntuar ROBOTRADER, la primera competición *algo trading* registrada en España. Si bien es cierto que el CR es calculado para tres años, una vez más la escasez de base científica no explica el por qué. Es por ello que nosotros aplicaremos un derivado para calcularlo así en otros rangos temporales y concluir los resultados obtenidos según los experimentos a simular. Podremos ver que al final de todo el proceso a estudiar y como comentábamos en el objetivo 3 del punto 1.6, lo más importante es reducir el riesgo y esto suele ser equivalente a reducir el M_{dq} . Por tanto cuanto mayor sea el CR, menos riesgo exponemos al inversor. Otra de las medidas de rendimiento para reducir el riesgo es por medio del consolidado estándar del llamado VaR.

Con el siguiente ejemplo podemos realizar los cálculos expresados arriba:

Supongamos que la curva de resultados (*equity*) o retorno obtenido en dólares por un experimento en base a un sistema que realiza un *trade* al día durante los meses de abril y mayo (*timeframe* de 60 minutos) es;

equity (\$) =

[1000000, 970000, 975000, 980000, 990000, 1000000, 1010000, 1060000, 1050000, 1045000, 1030000, 1020000, 1035000, 1010000, 1015000, 1020500, 1020000, 1017080, 1026503, 1030000, 1040000, 1039876, 1032878, 1039238, 1040000, 1049873, 1042384, 1050398, 1058395, 1052495, 1060000, 1069568, 1066345, 1068948, 1067654, 1066992, 1068234, 1064938, 1065298, 1069999, 1070000, 1071938, 1080000, 1050000, 1058989, 1057623, 1069876, 1067653, 1063486, 1069388, 1079348, 1078372, 1080000, 1067898, 1069999, 1078789, 1080000, 1087898, 1083456, 1086798, 1090000]

Podemos obtener fácilmente los *trades* ganadores y los *trades* perdedores a partir de la *equity* suponiendo que partimos con una cartera de 1000000 \$ y que los resultados son netos, contando con las comisiones y el posible *slippage* (término explicado en el siguiente punto). Se trata de un sistema que para cada *trade* (apertura + cierre, independientemente de si compró o vendió), abre la operación el día n para cerrarse la operación a una hora del día n+1, abrir una nueva operación y así sucesivamente. Por tanto hemos realizado 60 *trades* en los 61 días representados;

Trades ganadores (\$) =

[5000, 5000, 10000, 10000, 10000, 50000, 15000, 5000, 5500, 9423, 3497, 10000, 6360, 762, 9873, 8014, 7997, 7505, 9568, 2603, 1242, 360, 4701, 1, 1938, 8062, 8989, 12253, 5902, 9960, 1628, 2101, 8790, 1211, 7898, 3342, 3202]

Trades perdedores (\$) =

[30000, 10000, 5000, 15000, 10000, 25000, 500, 2920, 124, 6998, 7489, 5900, 3223, 1294, 662, 3296, 30000, 1366, 2223, 4167, 976, 12102, 4442]

La representación gráfica de la *equity* citada sería:

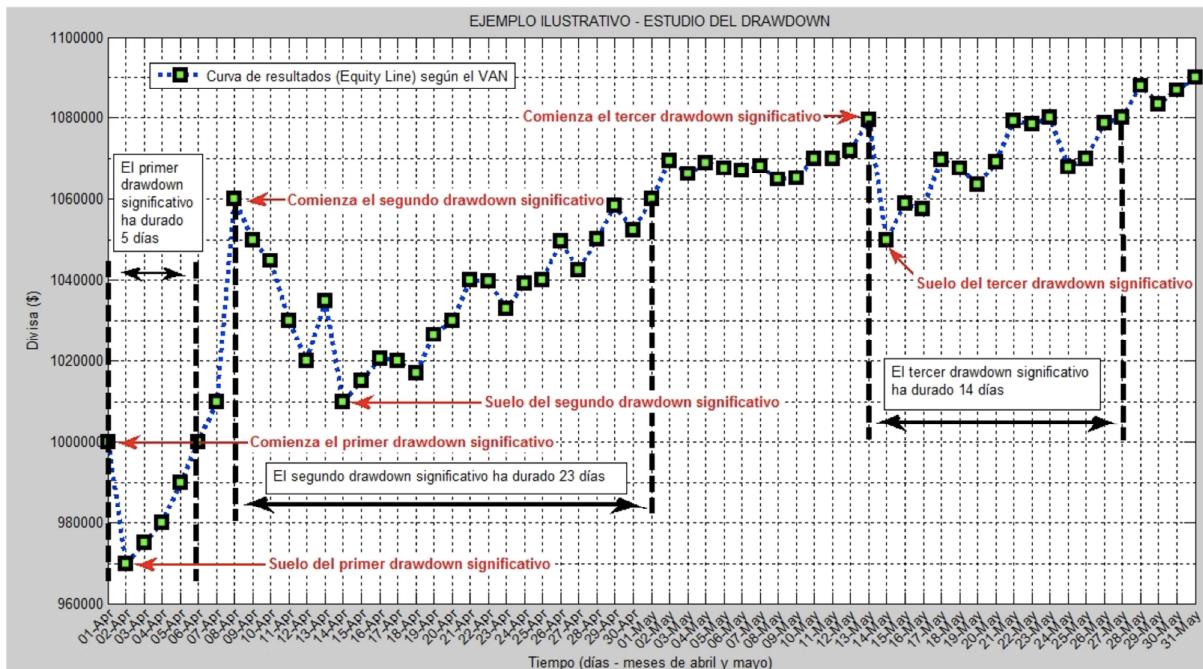


Figura 10: representación gráfica en MATLAB del retorno en el ejemplo del punto 2.4

Podemos apreciar varios *drawdowns* y en concreto tres candidatos significativos para ser el M_{dd} .

$$\text{Primer drawdown significativo} = 1000000\$ - 970000\$ = 30000 \$ (3\%)$$

$$\text{Segundo drawdown significativo} = 1060000\$ - 1010000\$ = 50000 \$ (4.72\%)$$

$$\text{Tercer drawdown significativo} = 1080000\$ - 1050000\$ = 30000 \$ (2.78\%)$$

En este caso $M_{dd} = \text{segundo drawdown significativo} = 30000 \$$

Es importante recordar que para estimar con mayor exactitud el M_{dd} en el análisis formal con los experimentos a realizar, utilizaremos el método Monte Carlo ya explicado. Si lo hubiésemos aplicado en el presente ejemplo el M_{dd} hubiera dado como resultado más de 30000 \$. Lo veremos más adelante.

$$\text{Por tanto, } CR = \frac{1090000\$ - 1000000\$}{30000\$} = 1.8$$

El resto de ratios dan como resultado:

$$SR(\text{normalizado}) = \frac{1500}{11262} (252 * 60 * 20/60) = 655.2$$

$$SQN = \frac{1500}{11262} * \sqrt{60} = 1.03$$

Pasamos ahora a exponer medidas de rendimiento computacionales cruciales para conocer si estamos o no perdiendo el tiempo y si tenemos la máquina idónea para los estudios realizados. Personalmente entiendo el *algo trading* como un conjunto de ligas. La liga de la inversión a largo plazo, medio plazo, corto plazo, intradiario y *High Frequency Trading*. ¿Qué liga deseas explotar? Antes yo me preguntaría, ¿Tengo los recursos técnicos para explotarla? Lógicamente otro factor importantísimo radica en la cartera a invertir porque uno de los estudios que en este trabajo no vamos a poder tratar en profundidad se basa en el análisis de la cartera mínima a invertir con un mercado o portfolio de ellos, con un sistema o portfolio de los mismos. Si dispongo de los recursos y cartera idóneos para entrar en la liga medio plazo por ejemplo, una operativa de 5-8 trades al mes y un retorno anual del 10% con una inversión inicial de 10000 euros. Hombre, lógicamente no dedicaré mi actividad laboral únicamente en el *algo trading*. Si opto por dedicarme por completo al *algo trading* y no dispongo de una amplia cartera, casi seguro necesitaré de terceros para lograrlo.

Volviendo a las medidas de rendimiento computacionales. Cuento con la ventaja de poder ejecutar los programas implementados en una máquina *Linux* y cito ventaja porque este sistema operativo, aparte de que es gratuito, es muy potente para obtener resultados sobre procesos y hebras ejecutándose en uno o varios cores (núcleos), entre otras funcionalidades. Lo normal es elegir una distribución *Linux* seguida por una amplia comunidad de desarrolladores, al menos eso es lo que yo hago, para que el soporte ante cualquier duda o problema casi seguro esté ya resuelta. Suelen ser las distribuciones más típicas *Ubuntu* y *Debian*, utilizaremos *Ubuntu*. Mediremos durante la optimización de parámetros, por experimento, lo siguiente:

- Tiempo de ejecución del programa.
- Carga media del sistema.
- Consumo de *RAM*.

2.5. Preprocesado de los históricos financieros

Los datos son el principal agente de la simulación, si no los consideramos fiables cómo vamos a estar satisfechos con los resultados obtenidos después del estudio realizado al Sistema Algorítmico de *Trading* o portfolio de ellos.

Como veníamos explicando, los datos financieros los podemos obtener de varias fuentes y éstas serán gratuitas o de pago. Dentro de las diferentes modalidades de pago nos transmitirán más o menos confianza. Pero ahora no toca sentir la confianza sino probarlo mediante unos filtros programados para identificar posibles irregularidades. En realidad vamos a encontrarnos problemas de dos principales tipos, político/empresarial y computacional. Aparte veremos los problemas particulares del mercado futuro, ya que cada producto financiero coteja sus propios problemas a considerar.

Por un lado, los problemas de tipo político/empresarial van a estar marcados principalmente por:

- Posibles cambios de horario en la contratación de un producto financiero en algún momento dado del histórico.
- Cambio de la moneda base, por ejemplo con los países que van adhiriéndose a Europa.
- Cambio del valor del *tick*, a lo mejor el futuro E-mini S&P 500 en unos años ya no será 0.25 y consideren cambiarlo. Suponiendo que siga existiendo dicho mercado.

Por si ocurre alguno de los puntos anteriores habrá que considerarlo cuando se estudia el sistema con el correspondiente histórico.

Por otro lado, la mayoría de las veces se trabaja con plataformas de trading propietarias y para las personas que no están familiarizados con técnicas de programación dan lugar a varios problemas básicos desde el punto de vista de la computación. Lo más probable es que el formato de los datos de una plataforma de *trading* propietaria (lenguaje de programación y estructura software programada propia) sea distinto del de otra plataforma propietaria. Es decir, si trabajamos por ejemplo con proveedor1 el fichero de datos deberá ser con el siguiente formato:

TICKER,PER,YYYYMMDD,TIME,OPEN,HIGH,LOW,CLOSE,VOLUME

simboloProducto,PER,AñoMesDía,hora,precioApertura,precioMáximo,precioMínimo, precioCierre,volumen

Nota: PER (Price Earning Ratio) es un ratio de tipo fundamental que tiene que ver con el mercado stock (acciones). En el mercado futuro no influye y la plataforma en este caso lo indica con una i latina mayúscula.

Mientras que si trabajamos por ejemplo con la plataforma proveedor2, el formato podrá ser variable según la cabecera que le pasemos con el fichero de datos. Es decir, un formato general podrá ser:

MM/DD/YYYY,HHMMSS,OPEN,HIGH,LOW,CLOSE,VOLUME

Mes/Día/Año,HorasMinutosSegundos,precioApertura,precioMáximo,precioMínimo, precioCierre,volumen

Fíjese que la base de datos en un fichero de texto (.txt) se puede implementar indicando espacio, coma, barra, punto y coma entre otros como separación entre los datos correspondientes. Se puede hacer también para la extensión .csv (representar datos en forma de tabla). Muchos de los lectores que conocen el arte de la computación, posiblemente usted, ya estarán pensando que es una barbaridad trabajar con un fichero .txt si vamos a abrirlo y cerrarlo por cada condición a analizar.

En efecto, yo mismo confirmo que es cierto, la instrucción `open()` y `close()` del lenguaje de programación C, aun siendo bajo nivel (cuanto más cerca del *hardware* programemos, más rápidos seremos), puede llegar a ser mortal si lo que se pretende es optimizar tiempo. La ventaja de realizar el presente estudio en MATLAB va a ser que, aun no siendo un entorno a tan bajo nivel como C, podemos guardarlos en un formato .mat y el mismo programa sólo tendrá que cargarlos una vez y guardarlos en una variable para procesarlos tantas veces como necesitemos. También podemos guardar el fichero en otras bases de datos, relacionales (MySQL, SQL) o no de tipo relacional (MongoDB, Hadoop, NoSQL), siendo éstas últimas las más estudiadas actualmente por la rapidez en las lecturas (procedimientos de *Map-Reduce* por ejemplo). Para ganar velocidad es muy importante reservar memoria previamente para las variables que necesitemos, lo más típico es inicializar las matrices a cero con la instrucción `zeros()`. Por ejemplo, si necesitamos guardar una muestra de 40000 combinaciones recorridos en un histórico de 70000 datos, tendremos entonces una variable que deberá guardar en RAM (como mínimo el programa a ejecutar) una matriz de $8*40000*70000 = 22.4$ GB. Si volvemos con el formato de los datos, la ventaja de conocer los lenguajes de programación es por ejemplo formatear cada línea del fichero para que lo pueda leer una plataforma concreta. Además, principalmente para los operadores discretionarios (operativa y/o identificación de patrones manualmente de forma visual entre otros) existen varias formas de representación gráfica de dichos datos. Por ejemplo está

el famoso gráfico candlestick (velas japonesas)

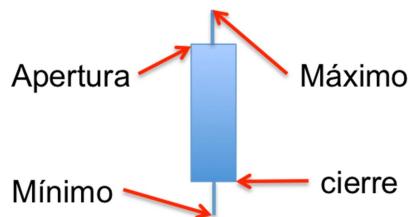


Figura 11: ejemplo representación tipo candlestick

de tipo barra

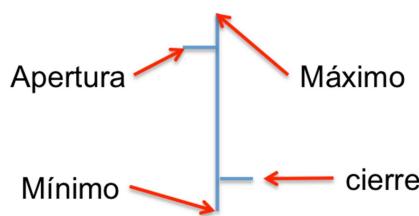


Figura 12: ejemplo representación tipo barra

y de tipo lineal (uniendo por ejemplo los precios de cierre) entre otros

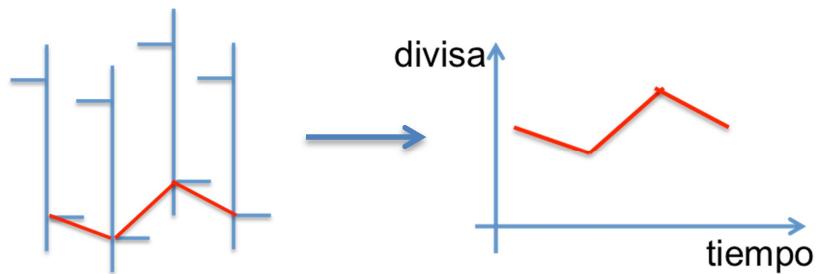


Figura 13: ejemplo representación tipo lineal con los precios de cierre

Fíjese que los datos básicos que utilizaremos no aportan todos los datos que intervienen en una ‘barra’ (sólo los extremos), si pagásemos por todos los datos necesarios del libro de órdenes (*bid, ask, price, position trades* y *arrivals trades* entre otros) podríamos decir que la simulación es perfecta porque contempla todas las condiciones del mercado en el pasado. Como no es así, lo normal en estos casos es penalizar los resultados obtenidos con el llamado *slippage*. Al igual que la comisión, el *slippage* contribuye, según una estimación, a la penalización con la divisa correspondiente sobre los supuestos beneficios conseguidos por el sistema en simulación. Generalmente se penaliza en puntos, mostraremos el resultado por el valor del *tick*. Por ejemplo en el E-mini S&P 500 tomaremos un *slippage* del $0.25 * \text{valorTick} = 0.25 * 12.5\$ = 3.125\$$. Si un sistema por ejemplo obtiene un retorno, contando comisiones, de 10000\\$ y realiza 100 trades, entonces el resultado estimado será $10000\$ - 100 * 3.125\$ = 9687.5\$$

A la pregunta, ¿Por qué 0.25 *ticks*? Como referencia, el regulador (CME Group) del mercado nos puede aportar el valor del *spread* (horquilla entre el *bid* y el *ask*) en su página web pero al final la mejor forma de conocer el verdadero *slippage* es, una vez más, partiendo de una buena cartera. Después de analizar el sistema, se supone que estamos bastante confiados por el procedimiento de estudio aplicado y además porque los resultados son favorables (el ratio beneficio/riesgo es alto). Entonces cogemos el sistema y con un solo contrato (lo que se aconseja hacer en el análisis previo a la gestión monetaria) lo ejecutamos con una cuenta real a mercado. Las operaciones que vaya realizando el sistema en tiempo real las guardamos y el mismo período de actuación en tiempo real lo pasamos por un *backtest* con el simulador. Se supone que el número de operaciones en *backtest* se debe corresponder con las operaciones en tiempo real (con la excepción de las operaciones de tipo ‘limitada’ o de ‘stop’) si se trata de órdenes lanzadas a mercado pero lo normal será que las operaciones de la simulación se ejecutan a un valor distinto que las operaciones en tiempo real. Pues bien, la media en la diferencia de ejecución en *ticks* de esas operaciones es el llamado *slippage*.

Como el presente trabajo está focalizado en el análisis cuantitativo de los datos, no nos importa su representación en el preprocesado de los datos (si acaso mostraremos algunas gráficas a modo de ejemplo).

Básicamente porque no es eficiente que me de la paliza de revisar más de un millón de líneas de una base de datos, comprobando una a una si los datos coinciden con otra fuente de datos o comparando gráfico a gráfico visualmente. Lo que si nos interesa es automatizar un procedimiento de filtrado para que en el caso de haber un posible error, el resultado del programa lo notifique.

Según lo explicado antes sobre el vencimiento de dicho futuro, tendremos un total de 16 vencimientos (2010-2013) para analizar por cada uno de los dos proveedores. En el punto 3 veremos en detalle los vencimientos seleccionados.

Como apunte importante a la hora de obtener los vencimientos en una plataforma de *trading*, se suele emplear un código de letras para identificar el mes.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto
Código	F	G	H	J	K	M	N	Q

Tabla 1: código vencimientos futuro de enero a agosto

	Septiembre	Octubre	Noviembre	Diciembre
Código	U	V	X	Z

Tabla 2: código vencimientos futuro de septiembre a diciembre

Por ejemplo, lo normal para pedir el vencimiento de septiembre de 2012 del futuro E-mini S&P 500, tendremos que indicar en el buscador de la plataforma del proveedor; ESU12 (códigomercadoCÓDIGOMESaño). Si en su lugar deseamos el continuo del mercado directamente teclearemos; ES. Pero éste último no nos interesa porque vamos a realizar un procedimiento para construir nuestro propio histórico continuo base de 1 minuto de *timeframe* (muestreando a continuación al *timeframe* de optimización) y así transmitir esa confianza de la que hablábamos anteriormente.

Como es lógico, dependiendo del nivel de detalle en los datos que tengamos, podremos realizar la comprensión que queramos. Es decir, si tenemos los datos con un *timeframe* de *tick* podremos obtener del mismo las comprensiones (muestreando, término archi-utilizado durante la carrera gracias a Nyquist) de segundos, minutos, días, semanas, meses y años, por supuesto. Pero no al revés, si contamos en el presente trabajo con un histórico de comprensión 1 minuto que nadie intente sacar en base a éste el histórico de *tick*. Ocurre lo mismo con el tamaño de una imagen, si tenemos por ejemplo un .jpg de 30x60 píxeles que nadie se empeñe en conseguir la misma calidad en 300x600 píxeles, donde no hay no se puede (salvo estimar). Se rellenarán, por muy avanzadas que estén las técnicas de reconstrucción, los huecos de píxeles aproximados y se notará, como se suele decir, que la imagen está pixelada.

Existen otros problemas que no dependen en cuanto a si la plataforma es propietaria o no. Más bien que el histórico de datos que disponemos en un fichero de una fuente particular son tantos números que no podemos estar seguros, al menos con una primera pasada visual de si son fiables o no.

Debemos automatizar un procedimiento para obtener los resultados computados. Volvemos con el proceso de estudio de los vencimientos, las notificaciones vendrán determinadas por los siguientes filtros:

- ¿Aparecen gaps (huecos en los datos)? ¿Cuántos? Gap: Línea con fecha y hora pero sin datos de cierre, apertura, máximo, mínimo y volumen. Lo normal es que si no hay negociación en un momento dado, se omita la información en el fichero pasando así al siguiente dato de negociación.
- Como tenemos los vencimientos en hora local España peninsular, ¿Aparecen datos algún sábado o domingo?
- ¿El vencimiento cierra a las 15:30 horas (España) el tercer viernes de cada mes?
- ¿El (máximo, mínimo, apertura, cierre) es > 0 ?
- ¿El máximo es \geq (apertura, cierre, mínimo)?
- ¿El mínimo es \leq (apertura, cierre)? | El mínimo ya está comprobado en la anterior condición.
- ¿El volumen es ≥ 0 ? Si tuviésemos un gráfico con *timeframe = tick* sería > 0 . Lógico, por cada *tick* que se produce ha resultado mínimo por un contrato vendido o comprado. Si es igual a 0 lo normal es que se omita como decía antes.
- ¿Cuántas líneas le corresponde a cada vencimiento por cada uno de los dos proveedores.

Después de las posibles notificaciones que pueda dar el programa toca la decisión de ajuste para suplir el posible error. Utilizaremos para ello algo parecido al ejemplo de la imagen que citábamos hace un momento. Aplicaremos un método de reconstrucción para que la estimación sea lo mejor posible. Realizaremos un ajuste con el método de mínimos cuadrados gracias a una curva polinómica sobreajustada apostando. Insisto, como voy a explicar en el punto 3.1.2, que la intención es preparar el histórico ya filtrado con una posible estimación ante posibles errores para tener confianza para comenzar el estudio y no de predecir como se supone debería hacer un Sistema Algorítmico de *Trading* en el mercado. Por ello que el polinomio se ajustará con el mayor exponente posible y así cuanto más ‘pegado’ esté la curva al precio, la estimación será más ajustada. Además, aunque los datos estén filtrados ante posibles errores no quiere decir que baste estudiar más filtros. Es decir, los datos podríamos decir que siguen estando ‘en crudo’ (en términos de la Telecomunicación, señal de información + señal de ruido) pero se podría aplicar la ecuación matemática idónea (por ejemplo un Filtro Paso Bajo es la media móvil simple), para filtrar la señal de ruido (en Telecomunicación; El ruido es la señal que enmascara la señal de información recibida en el receptor dificultando aquella la recuperación de ésta). Veámoslo en un ejemplo explicado lo más simple posible para entender a lo que me refiero:

Sea una señal sinusoide, en banda base, a una frecuencia f_m [Hz]:

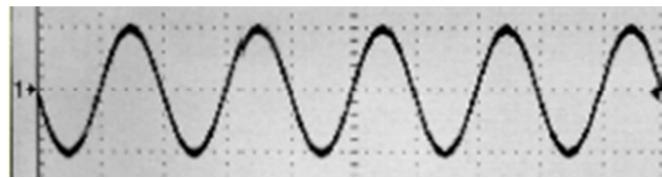


Figura 14: captura de los apuntes de la asignatura LECM (Laboratorio de Electrónica y Comunicaciones) de una señal moduladora en el dominio del tiempo

Trasladamos la señal en banda base (moduladora) a una frecuencia superior para su transmisión, aplicando por ejemplo la modulación AM, con una portadora de frecuencia f_c [KHz]:

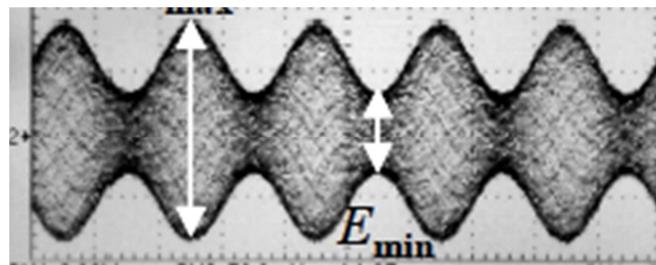


Figura 15: captura de los apuntes de la asignatura LECM (Laboratorio de Electrónica y Comunicaciones) de una señal modulada, AM, en el dominio del tiempo

La figura X muestra la señal que se va a transmitir al canal (alámbrico o inalámbrico), conteniendo en la misma la información, f_m . El receptor, una radio por ejemplo, dispone de un demodulador para que en función de la señal recibida pueda obtener la información, f_m . Pero, ¿Qué pasaría si se recibe en un receptor la señal a continuación?

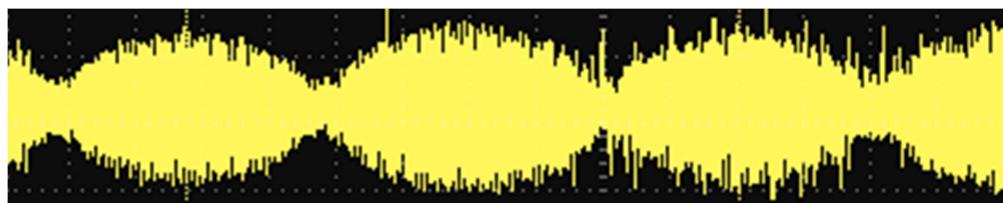


Figura 16: ejemplo señal modulada en AM con ruido en recepción en el dominio del tiempo

Ocurriría que un demodulador a lo mejor no podría obtener la información por el ruido que se ha producido durante su transmisión por el canal hasta llegar a recepción.

Por ello que, además de elegir una de las dos categorías de demoduladores (coherente y no coherente), se recomienda utilizar filtros (paso bajo y paso banda son los más empleados) en el circuito de transmisión y en el de recepción para reducir el ruido lo máximo posible y así obtener finalmente la información.

Añadir algún tipo de filtrado, como se necesita en la Telecomunicación, en una señal bursátil es una de mis tareas pendientes. Como puede ver, cualquier trabajo mejor hecho siempre se multiplica como es lógico.

Como curiosidad, La señal moduladora decíamos que tiene como información f_m , pues bien, en la Telecomunicación se aplica la Transformada de Fourier para conocer la frecuencia de una señal temporal. Es decir, la Transformada de Fourier de una sinusoides, $\cos(2\pi f_m t)$, es una delta simétrica:

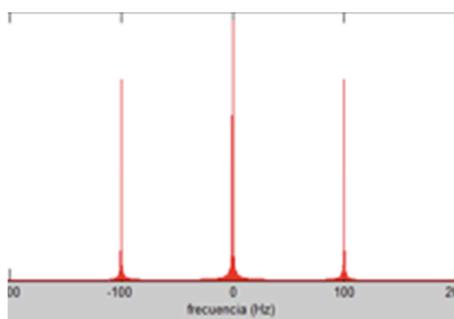


Figura 17: ejemplo señal moduladora, sinusoides, de frecuencia $f_m = 100$ Hz en el dominio de la frecuencia

A continuación la Transformada de Fourier del vencimiento de marzo del futuro E-mini S&P 500 (1 de diciembre de 2012 al tercer viernes de marzo de 2013):

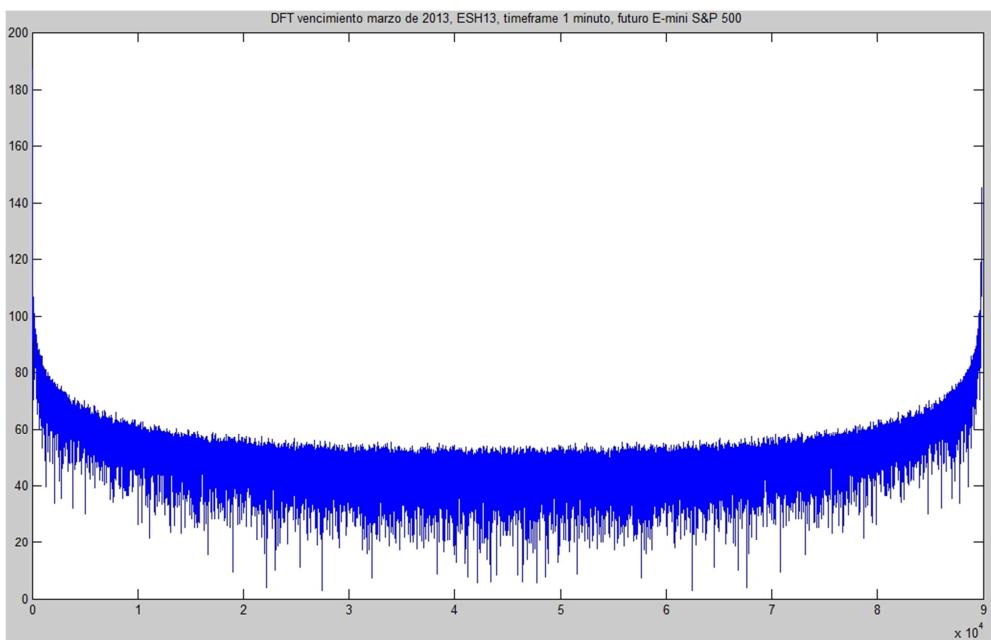


Figura 18: Transformada de Fourier del futuro E-mini S&P 500, vencimiento de marzo 2013, timeframe 1 minuto, fuente proveedor2

A continuación, la Transformada de Fourier del vencimiento de junio del futuro E-mini S&P 500 (1 de marzo de 2013 al tercer viernes de junio de 2013):

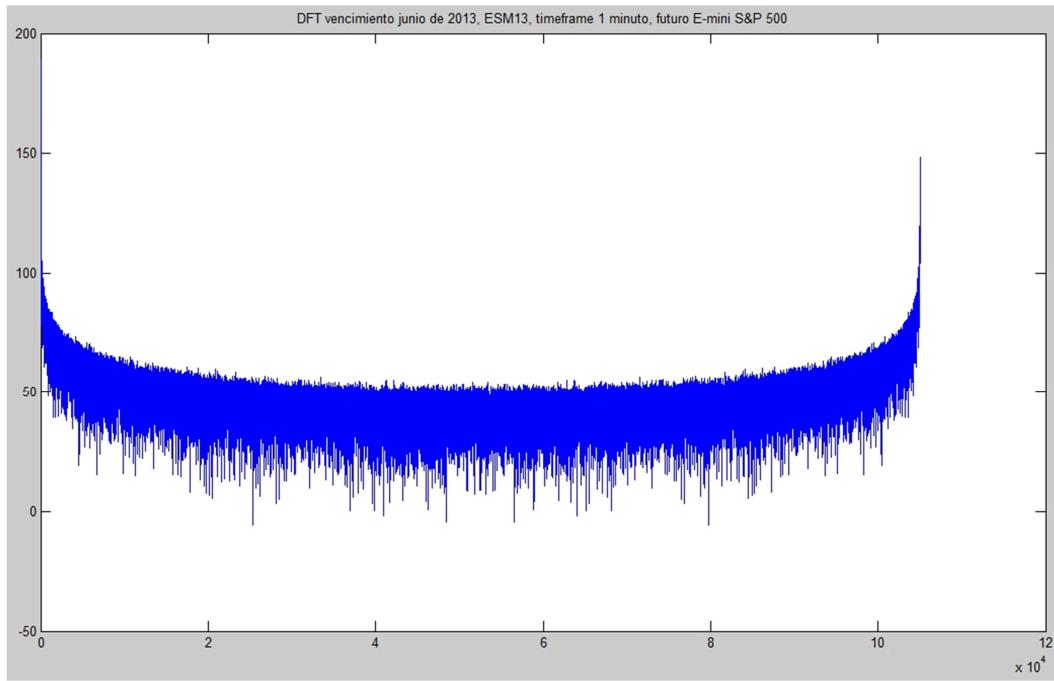


Figura 19: Transformada de Fourier del futuro E-mini S&P 500, vencimiento de junio 2013, timeframe 1 minuto, fuente proveedor2

A continuación la Transformada de Fourier del vencimiento de septiembre del futuro E-mini S&P 500 (1 de junio de 2013 al tercer viernes de septiembre de 2013):

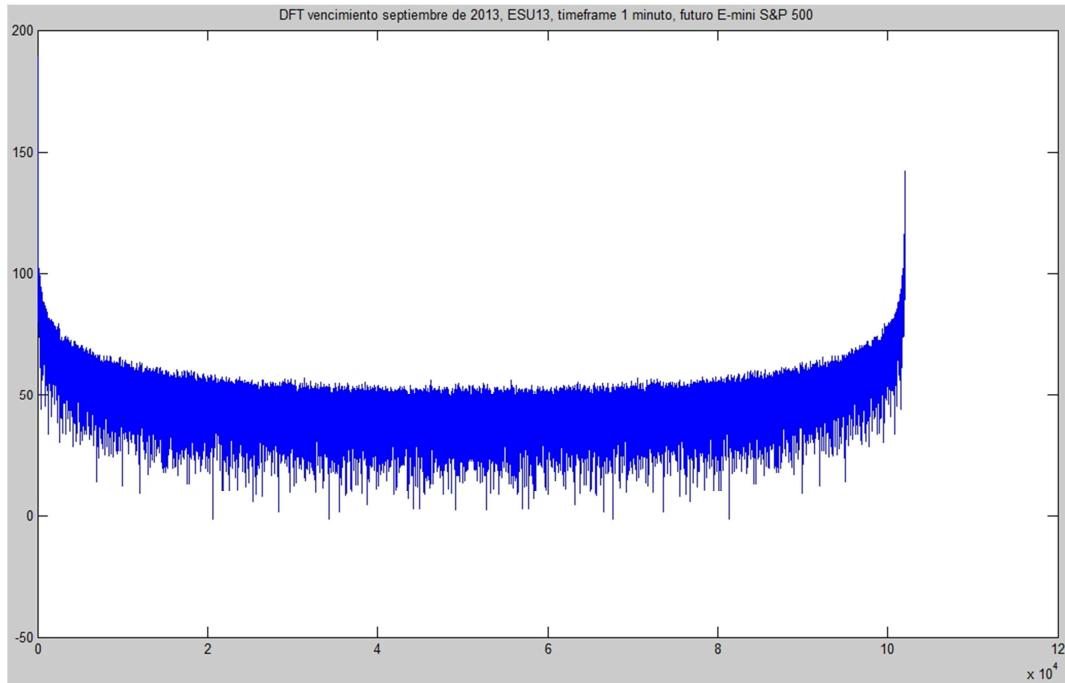


Figura 20: Transformada de Fourier del futuro E-mini S&P 500, vencimiento de septiembre 2013, timeframe 1 minuto, fuente proveedor2

A continuación la Transformada de Fourier del vencimiento de diciembre del futuro E-mini S&P 500 (1 de septiembre de 2013 al tercer viernes de diciembre de 2013):

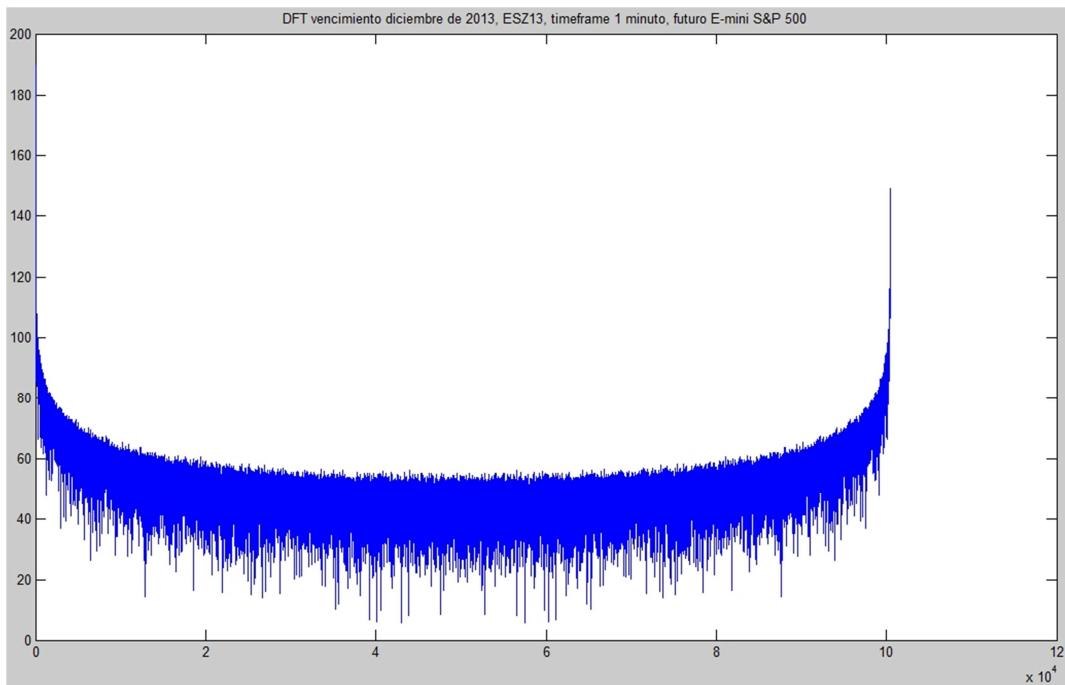


Figura 21: Transformada de Fourier del futuro E-mini S&P 500, vencimiento de diciembre 2013, timeframe 1 minuto, fuente proveedor2

Como se puede ver, no se aprecia en ningún vencimiento una frecuencia de mercado clara. Hay mucho ruido. Si en cambio, después de filtrar el ruido correctamente podemos obtener una delta (frecuencia de mercado), el indicador de tipo oscilador sería óptimo y la operativa espléndida.

Después de la anterior curiosidad volvemos con los filtros de los vencimientos. Sucesivamente pasaremos a la construcción del histórico continuo previo al estudio de optimización y de postprocesado de los datos resultantes de dicha optimización. Ya hemos explicado al comienzo del presente punto que surgen unos posibles problemas, previo a obtener el histórico financiero continuo, al tratarse de un mercado futuro y éstos los dividimos en dos;

1. Tenemos varias opciones para rolar (*rollover*) el vencimiento, las principales son (sin contar con procedimientos personalizados que pueda tener un operador bursátil con amplia experiencia):
 - a. Con fecha fija: tan simple como unir los vencimientos el tercer viernes de cada mes.

- b. Por cambio de volumen: Durante el presente vencimiento, la idea es analizar el volumen del vencimiento posterior y cuando éste sea superado por el presente vencimiento no esperamos al tercer viernes y aplicamos el rollover, sea el día que sea. Éste será el procedimiento que utilizaremos en los veinticuatro experimentos y que compararemos con los rendimientos del experimento número veinticinco y veintiséis (*rollover* con fecha fija). El volumen es la oferta y la demanda, por tanto si los operadores bursátiles están operando en el siguiente vencimiento antes del tercer viernes habrá que seguir a la oferta y la demanda para no quedarnos 'solos' en el mercado.
2. Disponemos de varios procedimientos de ajuste (*split*), los principales son (sin contar con procedimientos personalizados que pueda tener un operador bursátil con amplia experiencia):
- a. Supongamos que el vencimiento presente tiene el precio en 1000\$ pero cuando hacemos el rollover, el precio del siguiente vencimiento está a 1020\$, entonces cogemos TODO el histórico continuo que llevábamos y lo subimos 20\$.
 - b. Ratio Adjusted Data (RAD) [Thomas Stridsman]. Éste será el procedimiento que emplearemos, ¿por qué? Porque es un ajuste razonado a simple vista y es de los más utilizados. Pero para dar una respuesta más técnica tendría que elaborar otro Proyecto Fin de Carrera o quizás Tesis Doctoral para demostrarlo (bien avisé en el punto 1.5 que uno de los principales problemas se encuentra en la escasez científica existente para demostrar el por qué de todas las decisiones tomadas. Muchos estudios faltan por realizar, si no qué fácil sería ganar dinero en la bolsa.)
- $$RAD = cp * \left(1 + \frac{cf - cp}{cf}\right)$$

cp = precio cierre de la última barra en el vencimiento presente

cf = precio cierre de la barra en el momento del rollover (próximo vencimiento)

2.6. El problema de la sobreoptimización (*overfitting*)

Uno de los grandes problemas de la ciencia, cuando se trata de obtener resultados estadísticos/probabilísticos, radica en el razonamiento del procedimiento y escoger la apropiada muestra en su correspondiente población, para considerar los resultados fiables. Bien es cierto, como hemos reivindicado varias veces hasta el momento, que gracias a los avances en la tecnología computacional, las máquinas cada vez son más rápidas y con mayor capacidad de memoria para analizar una gran serie de datos. Por ende, se supone que con un mayor análisis, la seguridad aumentará. Pero ¿es correcto cómo se analiza? Las herramientas habituales, como ya hemos visto, son una de tantas como pueda haber actualmente en el mundo. El problema radica en razonar cómo se va a proceder de forma justificada. Los algoritmos de aprendizaje necesitan un entrenamiento para ajustar los parámetros del problema a resolver y a continuación, con los parámetros ajustados, estimar en mejor o menor medida la predicción del futuro. Pues bien, en el problema que nos acomete, la optimización no es un algoritmo de aprendizaje* estrictamente hablando pero se parece en el procedimiento que pretendemos realizar.

2.6.1. Métodos de medida

Ya hemos hablado varias veces, llegados a este punto, en términos de ‘muestra’ y ‘población’. Términos estrechamente ligados con la estadística. Según el Instituto Nacional de Estadística de España, la palabra estadística tiene dos acepciones en general, por un lado es el hecho de estudiar las características de una población y sus integrantes; y por otro lado es una disciplina científica que entre muchas otras utilidades puede usarse para deducir relaciones entre variables, o para extender los resultados que obtengamos para una parte de la población a toda la población. Ésto último es importante “...para extender los resultados que obtengamos para una parte de la población (muestra) a toda la población...”. Si por ejemplo deseamos realizar un estudio para estimar la intención de voto de un país, ¿preguntaremos a toda la población del país? No creo que sea barato realizar dicho estudio, de hecho el Instituto Nacional de Estadística no creo que considere presupuestarlo de esa manera, irónicamente hablando. Lo mismo ocurre con los datos del mercado financiero, ¿Vamos a proceder a estudiar los parámetros de un Sistema Algorítmico de *Trading* en toda la población del mercado desde su fundación? No, seleccionaremos para ello una muestra y lógicamente existirá un error por no considerar todos los casos. De la muestra seleccionada, lo normal en la comunidad científica, dividiremos una parte para entrenar y la otra para probar como si de un supuesto futuro se tratase. Al igual que un algoritmo de aprendizaje, como comentábamos antes, primero entrenamos (optimizar) con la muestra IS del histórico financiero para ajustar los parámetros (un 75% por ejemplo) y con los parámetros ajustados probamos los resultados, *backtest*, en la otra parte de la muestra, OOS (el 25% restante).

*Weka, de la universidad de Waikato, ya trae implementado en JAVA muchos algoritmos de aprendizaje.

Cabe destacar que existen dos bandos en la comunidad científica, todavía queda mucho por demostrar, unos postulan que los resultados OOS obtenidos son suficientes pero otros se mantienen en que una muestra OOS sigue siendo el pasado a pesar de tratarlo como un supuesto futuro y por ello que un procedimiento Monte Carlo para tratar una muestra aleatoria es más recomendable. Bien, primer paso conseguido, ya hemos seleccionado la muestra del histórico financiero.

El segundo paso se basa en los parámetros a entrenar del Sistema Algorítmico de Trading en la muestra ya seleccionada del primer paso. Entonces, ¿Probaremos todas las posibles combinaciones de los parámetros del Sistema Algorítmico de *Trading*? Por lógica, algún parámetro supongo que pertenecerá al conjunto de los números reales, \mathbb{R} (ya sean naturales, \mathbb{N} , enteros, \mathbb{Z} , racionales, \mathbb{Q} o irracionales, \mathbb{I}), por tanto dígame qué máquina va a utilizar para que en un período razonable de tiempo los pueda optimizar todos y me dice el modelo para comprármela yo también. Lo normal, ya de entrada, será acotar con un intervalo el rango de posibles valores de los infinitos posibles o hasta un máximo de lo que te permita la muestra elegida del primer paso. Acabamos de ver que vamos a tener otra muestra, esta vez en función de todos los posibles valores que pueda tomar un parámetro. Ya tenemos las dos muestras seleccionadas, ¿Ahora qué? Paso 3, ¿los resultados están sobreoptimizados?

Nota: El conjunto de los números reales, \mathbb{R} , tiene infinitos números. Por ejemplo el conjunto $(1,2)$.

El tercer paso es comprobar que los resultados no están sobreoptimizados. Antes de ver los distintos procedimientos, tenemos que ser conscientes que según las características del ordenador disponible, éste podrá o no procesar los resultados debido a los cálculos que va a tener que hacer cada combinación de parámetros con la otra muestra seleccionada del histórico financiero. Llegados a este momento se debería conocer por algún procedimiento si el proceso está o no sobreoptimizado. El inconveniente es que no se suele seguir un procedimiento técnico para concluir. En estos casos se suele valorar subjetivamente sobre los resultados obtenidos IS y OOS en función del ratio escogido. Es decir, si el ratio OOS no baja más de un porcentaje “desde mi punto de vista” respecto al ratio obtenido IS entonces consideramos que no está sobreoptimizado. Es cierto, como si de un axioma se tratase, dado un rendimiento IS muy superior al rendimiento OOS entonces suele ocurrir que los resultados en una supuesta operativa real pueden ser desastrosos y ésto es debido a la sobreoptimización. Por ello que un rendimiento IS parecido al rendimiento OOS se suele decir que es robusto, consistente y por tanto no sufre sobreoptimización. También existen procedimientos derivados, para hablar de sobreoptimización, de la base de la muestra tomada y dividida en IS y OOS. Un derivado de ellos es el llamado *Walk Forward Optimization (WFO)*, se trata de optimizar un período dado y dentro de dicho período dividir en varios IS y OOS para que a lo largo de la optimización se pueda comprobar una evolución de la posible robustez del sistema. Un ejemplo de ello viene a continuación:

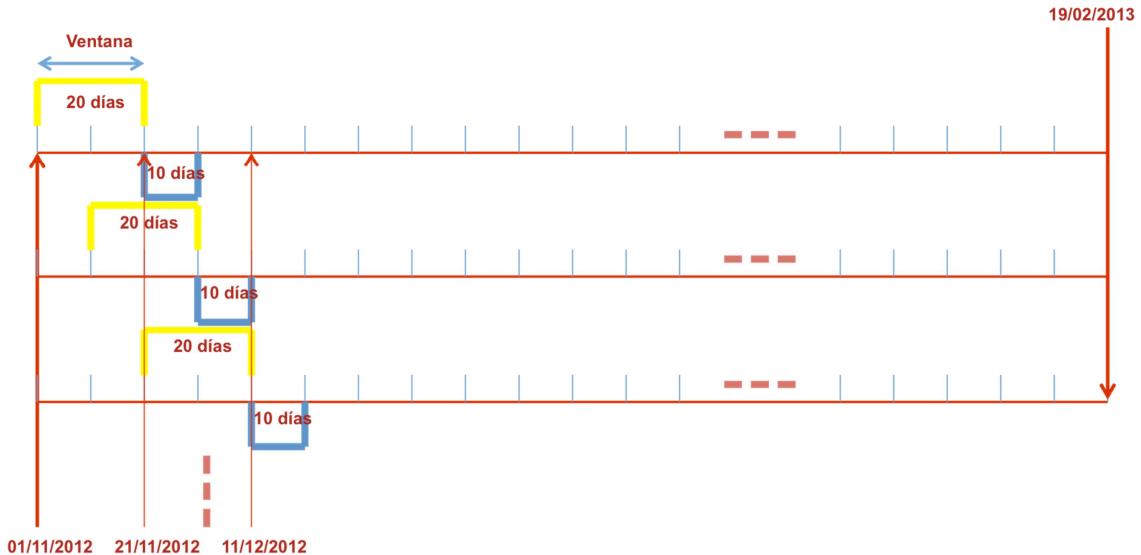


Figura 22: ejemplo del procedimiento Walk Forward Optimization

Como se puede observar, el WFO de la figura X está optimizando el período del histórico desde el 1 de noviembre del 2012 hasta el 19 de febrero del 2013. Se ha seleccionado en este caso una ventana de período de optimización de 20 días (IS) y la mejor combinación de parámetros aplicarlo a los 10 días siguiente (OOS). Desplazamos la ventana temporal IS 10 días y obtenemos los nuevos resultados. Así sucesivamente hasta que llegamos al final. Los resultados parciales los obtenemos y podremos así comprobar cómo se hubiera comportado el sistema en todo el período. En mi opinión es un procedimiento interesante para aplicar pero como inconvenientes a tomar en cuenta considero dos; 1. Las ventanas son fijas de principio a fin, faltaría estudiar un caso con ventanas dinámicas y 2. ¿Cómo obtiene la mejor combinación de parámetros el procedimiento WFO? Si lo que ejecuta una plataforma concreta es simplemente escoger el primero de todos los resultados optimizados sin realizar como mínimo el estudio que estamos presentando en este trabajo, quizás los resultados estén bastante distorsionados debido a casualidades estadísticas.

Entonces ¿finalmente no podremos obtener un supuesto resultado objetivo de sobreoptimización? La respuesta es si, gracias a los estudios realizados por;

- David H. Bailey, Investigador del Departamento de Computer Science en la Universidad de California.
- Jonathan M. Borwein, Profesor de Matemáticas en la Universidad de Newcastle.
- Marcos López de Prado, Responsable del Área Cuantitativa de Trading en Hess Energy Trading Company (HETC).
- Qiji Jim Zhu, Profesor de Matemáticas en la Universidad de Western Michigan.

Vamos a ver en el siguiente punto un procedimiento científico, publicado actualmente como *Abstract*, para obtener la probabilidad de sobreoptimización (PBO). Veremos que los resultados tienen algo que ver con las sentencias subjetivas ya mencionadas.

Al tratarse de un *Abstract*, todavía queda mucha investigación por delante y el presente trabajo participa aportando resultados necesarios para dicha investigación.

2.6.2. Cuantificación de la sobreoptimización

En base a la muestra del histórico financiero explicada ya en el punto anterior, es importante considerar que analizaremos tres años, 2010-2012, de IS (con el histórico continuo construido por nosotros mismos) y un año de testeo o *backtest*, 2013, a razón del OOS. A la pregunta, ¿Por qué tres años de IS y uno de OOS? Una vez más, existe escasez científica con una demostración objetiva para ello. En cualquier rama científica, generalmente se escoge una muestra mayor de IS que de OOS y dichas decisiones según unos criterios subjetivos adquiridos por la experiencia. Se habla subjetivamente, pero con lógica, que la muestra debe contener todas las posibles situaciones del pasado y éstas lo más probable es que vuelvan a ocurrir en el futuro. En el área que ahora nos corresponde, financiero, podríamos justificar que la mejor muestra para analizar será aquella que comprenda situaciones de tendencia alcista, bajista y lateral con baja y alta volatilidad. Es decir, seis posibles situaciones que suelen categorizar los desarrolladores de sistemas. Una tarea personal que tengo pendiente es poder automatizar un procedimiento para obtener dichas seis situaciones porque aun optando por tres años y suponer que ya las contempla, lo que más nos interesa es optimizar el tiempo. Es decir, si el procedimiento automatizado concluye que es suficiente con dos años de histórico para IS, no estamos ahorrando tiempo de cómputo al analizar el tercer año aunque no vaya a aportar más información de la necesaria. Para lo que si puede ayudarnos tomar una año más de histórico, dependiendo del Sistema Algorítmico de Trading, es en cuanto al número de operaciones. Para considerar una muestra significativa por si nos quedamos cortos en número de *trades* con dos años de histórico. Además de la muestra seleccionada a razón del histórico financiero, realizaremos cuatro experimentos, explicados en detalle en el punto 3.1.1, variando el número de combinaciones según la segunda muestra que debemos considerar a razón de los parámetros del sistema. Por último, consideraremos dos líneas de investigación con dos medidas de rendimiento como función *fitness* a analizar (explicado en el punto 3.1.1).

A continuación veamos el procedimiento técnico en cinco pasos, acompañado con un ejemplo, elaborado por los autores citados al final del punto anterior y que será la base de estudio en referencia a la sobreoptimización (PBO: ProBabilidad de Overfitting).

Paso 1:

Creamos una matriz, M_{TxN}

N = última combinación de parámetros ejecutada en optimización (IS). Cada columna es el vector P&L según los *trades* realizados, contando posibles *trades* abiertos.

Ejemplo:

Tenemos un sistema con tres parámetros y deseamos realizar una optimización (ya sea ejecución lineal o por algoritmos genéticos) maximizando el Ratio SQN (función *fitness*), tomando un histórico financiero (3 años) y *timeframe* (15 minutos) fijo y con las siguiente combinación de parámetros:

[valor inicial : salto : valor final]

Parámetro1 [1 : 2 : 50]

$$\frac{|50 - 1|}{2} + 1 = 25.5 \{truncamos\} 25$$

Parámetro2 [0.1 : 1 : 0.9]

{multiplicamos por 10 hasta que desaparezcan los decimales} [1 : 1 : 9]

$$\frac{|9 - 1|}{1} + 1 = 9$$

Parámetro3 [-8 : 2 : -6]

$$\frac{| -8 + 6 |}{2} + 1 = 2$$

Tenemos un total de $25 * 9 * 2 = 450$ combinaciones para optimizar. Por tanto:

$$M_{Tx450} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,450} \\ \dots & \dots & \dots & \dots \\ x_{T,1} & x_{T,2} & \dots & x_{T,450} \end{bmatrix}$$

Cada columna representa, como ya hemos dicho, un vector P&L por cada combinación de parámetros dada y una función de *fitness*.

Paso 2:

Partimos M_{TxN} en filas, formando así submatrices, $S_{\frac{T}{s}xN}$ y cada una de ellas con la misma dimensión.

s = número par

Al igual que ocurre con los posibles parámetros que cuenta el algoritmo genético, añadimos, en el procedimiento que estamos explicando para obtener la probabilidad de *overfitting*, un parámetro. Realizo las pruebas con el parámetro s a 10.

Hay que considerar que sólo un Proyecto Fin de Carrera o quizás Tesis Doctoral podría enfocarse en el testeo y/o mejora del procedimiento que estamos explicando, optimizando el parámetro s para obtener una mayor sensibilidad de la probabilidad PBO por ejemplo, incluso añadir más posibles parámetros a este procedimiento tan reciente en el campo del trading (agosto de 2013). Cito en el campo del *trading* porque es el que estamos tratando en el trabajo, pero se podría estudiar el mismo procedimiento en otra área, como la robótica.

Continuamos con el ejemplo:

$$s = 4 \quad \rightarrow \quad M_{TxN} \begin{bmatrix} sector1 \\ \dots \\ sector2 \\ \dots \\ sector3 \\ \dots \\ sector4 \end{bmatrix}$$

Cada sector está formado por una submatriz de dimensiones $\frac{T}{4} \times 450$

Paso 3:

Sacamos todas las combinaciones posibles sin repetición de submatrices (tomando grupos de $\frac{s}{2}$). Se trata de combinatoria sin repetición porque el resultado de la función *fitness* es el mismo que en el caso de combinatoria con repetición.

Continuamos con el ejemplo:

$$\binom{\frac{s}{2}}{s/2} = \binom{4}{2} = \frac{4!}{(4-2)! * 2!} = 6$$

$$S_1 = \begin{bmatrix} sector1 \\ \dots \\ sector2 \end{bmatrix} \text{ dimensión } \frac{T}{2} \times 450 \qquad S_2 = \begin{bmatrix} sector1 \\ \dots \\ sector3 \end{bmatrix} \text{ dimensión } \frac{T}{2} \times 450$$

$$S_3 = \begin{bmatrix} sector1 \\ \dots \\ sector4 \end{bmatrix} \text{ dimensión } \frac{T}{2} \times 450 \qquad S_4 = \begin{bmatrix} sector2 \\ \dots \\ sector3 \end{bmatrix} \text{ dimensión } \frac{T}{2} \times 450$$

$$S_5 = \begin{bmatrix} sector2 \\ \dots \\ sector4 \end{bmatrix} \text{ dimensión } \frac{T}{2} \times 450 \qquad S_6 = \begin{bmatrix} sector3 \\ \dots \\ sector4 \end{bmatrix} \text{ dimensión } \frac{T}{2} \times 450$$

Paso 4:

Construir el vector $\overline{R_{ISk}}$ ($k = 1, 2, \dots, s$) con los resultados de la función fitness, In Sample (IS). Entrenamiento.

Construir el vector $\overline{R_{OOSk}}$ ($k = 1, 2, \dots, s$) con los resultados de la función fitness, Out Of Sample (OOS). Testeo.

Para ello cogemos una submatriz para entrenamiento (IS) y por otro lado buscamos la submatriz complemento para el testeo (OOS).

Es decir, si continuamos con el ejemplo:

$$J_{IS1} = \begin{bmatrix} \text{sector1} \\ \dots \\ \text{sector2} \end{bmatrix} \quad J_{OOS1} = \begin{bmatrix} \text{sector3} \\ \dots \\ \text{sector4} \end{bmatrix}$$

A razón de J_{IS} tendremos un resultado de la función *fitness* (ff) por columna y así obtener:

$$\overline{R_{IS1}} = (ff_{IS1}, ff_{IS2}, \dots, ff_{ISN})$$

A razón de J_{OOS} tendremos un resultado de la función *fitness* (ff) por columna y así obtener:

$$\overline{R_{OOS1}} = (ff_{OOS1}, ff_{OOS2}, \dots, ff_{OOSN})$$

El mismo proceso explicado lo realizamos con todas las parejas de submatrices, en este ejemplo 6 grupos:

$$\begin{array}{ll} J_{IS1} = \begin{bmatrix} \text{sector1} \\ \dots \\ \text{sector2} \end{bmatrix} & J_{OOS1} = \begin{bmatrix} \text{sector3} \\ \dots \\ \text{sector4} \end{bmatrix} \\ J_{IS2} = \begin{bmatrix} \text{sector3} \\ \dots \\ \text{sector4} \end{bmatrix} & J_{OOS2} = \begin{bmatrix} \text{sector1} \\ \dots \\ \text{sector2} \end{bmatrix} \\ J_{IS3} = \begin{bmatrix} \text{sector1} \\ \dots \\ \text{sector3} \end{bmatrix} & J_{OOS3} = \begin{bmatrix} \text{sector2} \\ \dots \\ \text{sector4} \end{bmatrix} \\ J_{IS4} = \begin{bmatrix} \text{sector2} \\ \dots \\ \text{sector4} \end{bmatrix} & J_{OOS4} = \begin{bmatrix} \text{sector1} \\ \dots \\ \text{sector3} \end{bmatrix} \end{array}$$

$$J_{IS5} = \begin{bmatrix} sector1 \\ \dots \\ sector4 \end{bmatrix} \quad J_{OOS5} = \begin{bmatrix} sector2 \\ \dots \\ sector3 \end{bmatrix}$$

$$J_{IS6} = \begin{bmatrix} sector2 \\ \dots \\ sector3 \end{bmatrix} \quad J_{OOS6} = \begin{bmatrix} sector1 \\ \dots \\ sector4 \end{bmatrix}$$

Después de obtener los vectores $\overline{R_{IS}}$ y $\overline{R_{OOS}}$ de todas las parejas de submatrices, procedemos a realizar unos cálculos con dichos vectores.

Dado $\overline{R_{ISk}} = (ff_{IS1}, ff_{IS2}, \dots, ff_{ISN})$ debemos calcular el elemento n^* :

$$n^* = \arg \max_n \{ff_n\}$$

Es decir, cuál es el mayor resultado de la función *fitness* del vector $\overline{R_{ISk}}$ ($k = 1, 2, \dots, s$) y anotamos la columna donde se encuentra n^* .

A continuación obtenemos el elemento w^* , asignándole el resultado del vector $\overline{R_{OOSk}}$ en la misma columna donde se encontraba n^* en el vector $\overline{R_{ISk}}$. Después el vector $\overline{R_{OOSk}}$ lo normalizamos ($\overline{\overline{R_{OOSk}}}$) y ordenamos de 0 a 1 para obtener el elemento w^* normalizado ($\overline{\overline{w^*}}$). Finalmente calculamos λ_c :

$$\lambda_c = \ln \frac{\overline{\overline{w^*}}}{1 - \overline{\overline{w^*}}} \in (-\infty, \infty)$$

En el área del *trading* lo normal será que la función *fitness* tenga como objeto maximizarla (aunque existen funciones *fitness* que su objetivo es minimizarla) y por tanto desearemos una λ_c que tienda al infinito. Si $\lambda_c = 0$ querrá decir que coincide con la mediana (Me) del vector $\overline{\overline{R_{OOSk}}}$.

Si derivamos el ejemplo que estábamos siguiendo, ahora suponemos que $N = 6$ y no 450.

$$\overline{R_{IS1}} = (1.1, 3, 0.9, 3.1, 1.5, 1.7) \quad \rightarrow \quad n^* = 3.1 \text{ (cuarta columna)}$$

$$\overline{R_{OOS1}} = (0.5, 0.9, 1.2, 0.1, 0, 2) \quad \rightarrow \quad w^* = 0.1 \text{ (cuarta columna)}$$

Normalizamos $\overline{R_{OOS1}}$. Es decir, escogemos dos puntos (por ejemplo el mínimo y el máximo) y creamos ecuación de la recta. Este procedimiento de normalizar nos recuerda al utilizado en el comienzo del desarrollo del algoritmo genético.

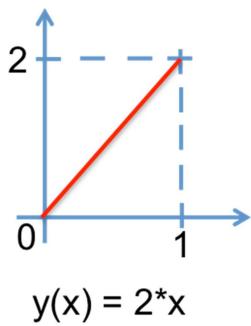


Figura 23: normalización del vector $\overline{\overline{R_{OOS1}}}$

$$\overline{\overline{R_{OOS1}}} = (0, 0.05, 0.25, 0.45, 0.6, 1)$$

$$w^* = 0.1 \text{ (cuarta columna)} \quad \rightarrow \quad \overline{\overline{w^*}} = 0.05 \quad \rightarrow \quad \lambda_c = \ln \frac{0.05}{1 - 0.05} = -2.94$$

Los λ_c de las seis parejas de submatrices nos permitirá realizar una representación gráfica de la distribución de frecuencias y lo normal será observar una forma de campana de Gauss. La verdad es que gracias a las propiedades de simetría, función densidad, media y varianza entre otros, la distribución normal es muy requerida para modelar fenómenos. En la Telecomunicación sin ir más lejos se empleo por ejemplo como elemento decisivo de una sistema de comunicación digital que recibe unos y ceros.

Paso 5:

Cálculo de la probabilidad de overfitting (PBO).

$$PBO = p(w^* < Me[\overline{\overline{R_{OOSk}}}]) \quad k = 1, 2, \dots, s$$

Terminando con el ejemplo:

Recordamos que hemos supuesto al final $N = 6$ y $s = 4$, tenemos por tanto 6 submatrices y por tanto 6 parejas (IS/OOS). Es decir, 6 resultados de λ_c . Suponemos que las parejas 3 ($\overline{\overline{R_{IS3}} / \overline{\overline{R_{OOS3}}}}$), 4 ($\overline{\overline{R_{IS4}} / \overline{\overline{R_{OOS4}}}}$), 5 ($\overline{\overline{R_{IS5}} / \overline{\overline{R_{OOS5}}}}$) y 6 ($\overline{\overline{R_{IS6}} / \overline{\overline{R_{OOS6}}}}$) no tienen w^* por debajo de la mediana.

$$\overline{\overline{R_{IS1}}} = (1.1, 3, 0.9, 3.1, 1.5, 1.7) \quad \rightarrow \quad n^* = 3.1 \text{ (cuarta columna)}$$

$$\overline{\overline{R_{OOS1}}} = (0.5, 0.9, 1.2, 0.1, 0, 2) \quad \rightarrow \quad w^* = 0.1 \text{ (cuarta columna)}$$

$$\text{Si ordenamos } \overline{\overline{R_{OOS1}}} \rightarrow \widetilde{R_{OOS1}} = (0, 0.1, 0.5, 0.9, 1.2, 2) \quad \rightarrow Me = \frac{0.5 + 0.9}{2} = 0.7$$

En este caso, w^* si es menor a la mediana.

$$\overline{R_{IS2}} = (1, 1.2, 2, 0.2, 0.5, 1.1) \rightarrow n^* = 2 \text{ (tercera columna)}$$

$$\overline{R_{OOS2}} = (0.9, 1.4, 1.3, 1.9, 2, 0.1) \rightarrow w^* = 1.3 \text{ (tercera columna)}$$

$$\text{Si ordenamos } \overline{R_{OOS2}} \rightarrow \widetilde{R_{OOS2}} = (0.1, 0.9, 1.3, 1.4, 1.9, 2) \rightarrow Me = \frac{1.3 + 1.4}{2} = 1.35$$

En este caso, w^* si es menor a la mediana.

$$PBO = \frac{\text{casos favorables}}{\text{casos posibles}} = \frac{2}{6} = 0.33 \text{ (33%)}$$

Me parece que una pregunta que se puede estar haciendo en este momento es, a pesar de que vayamos a demostrar que a medida que PBO aumenta, el rendimiento OOS disminuye, ¿Dónde se encuentra la frontera de decisión? ¿A partir de qué probabilidad sufrimos un riesgo preocupante de sobreoptimización? Pues bien, bienvenido al increíble mundo de la estimación, otro parámetro que se podría estudiar en otro trabajo más específico o quizás tesis doctoral y que en cierta parte lo veremos más o menos con los resultados de los experimentos que realizaremos por cada una de las dos funciones *fitness* y filtro por número de *trades*.

El procedimiento que hemos visto lo representaremos, para referirnos a él, con el símbolo:

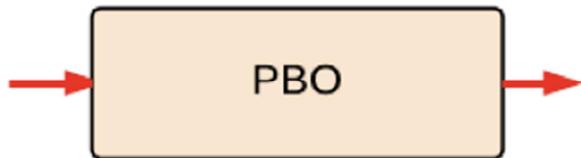


Figura 24: símbolo para referenciar el procedimiento de cálculo de la probabilidad de overfitting (sobreoptimización)

3. Metodología

Para demostrar los objetivos expresados en el punto 1.6 procederemos con un proceso de análisis dividido en tres fases. La fase 1 está orientado al preprocesado de los datos de los mercados financieros, la fase 2 (buscamos robustez) está orientado a la optimización (posibles filtros añadidos), el cálculo de la probabilidad de sobreoptimización (overfitting) y selección de la mejor combinación de parámetros y finalmente en la fase 3 (estudio del riesgo) estudiaremos el máximo *drawdown* con un método de Monte Carlo y el rendimiento beneficio/riesgo con el *Ratio Calmar*. Entre los distintos elementos a estudiar del conjunto del proceso veremos resultados parciales de evaluación. A dicho proceso lo voy a bautizar con el nombre ‘Proceso ATurel’. Ciento es que todas las herramientas que empleo ya han sido inventadas, que además, seguro existen procesos de análisis que se le parecerán más o menos pero al igual que ocurre con una novela, toda ella salió del diccionario ya inventado. Detallaremos en los siguientes puntos el estudio de cada elemento y englobaremos el proceso completo en un diagrama de flujos para su rápida comprensión.

3.1. Metodología Seguida

Antes de pasar a explicar cada elemento funcional (cajas rectangulares) del Proceso ATurel, a continuación podemos ver el Proceso ATurel completo representado en un diagrama de flujos de forma resumida.

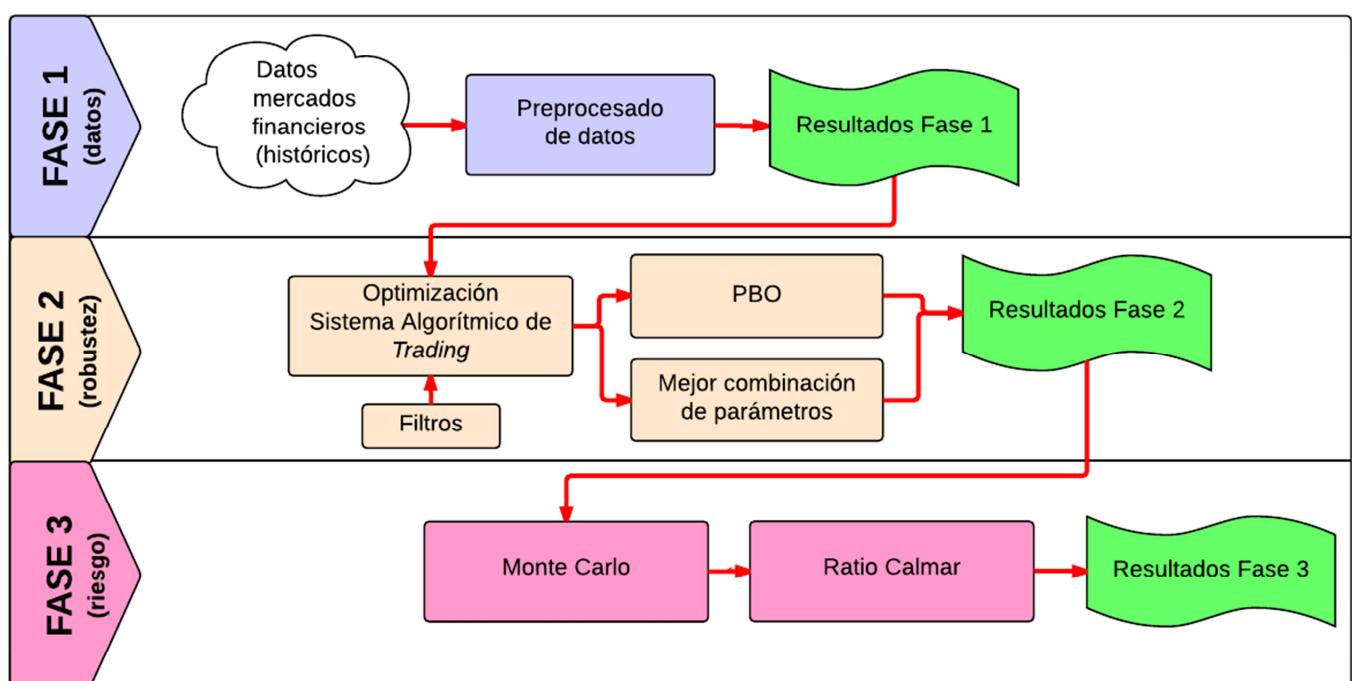


Figura 25: Proceso ATurel

Pasamos entonces a explicar el estudio realizado por cada uno de los elementos funcionales (cajas rectangulares) que forman el proceso ATurel.

Comenzamos con la fase 1,

‘Preprocesado de datos’:

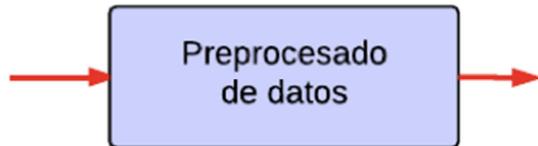


Figura 26: preprocesado de datos - Proceso ATurel

Los vencimientos a preprocesar del futuro E-mini S&P 500 están en hora local, España peninsular. En este trabajo estamos realizando el estudio con dos proveedores, lógicamente se podría aplicar el preprocesado, si se dispone, a más proveedores. Vamos a dividir el estudio en tres partes;

- **Filtrado 1 (borrado)**

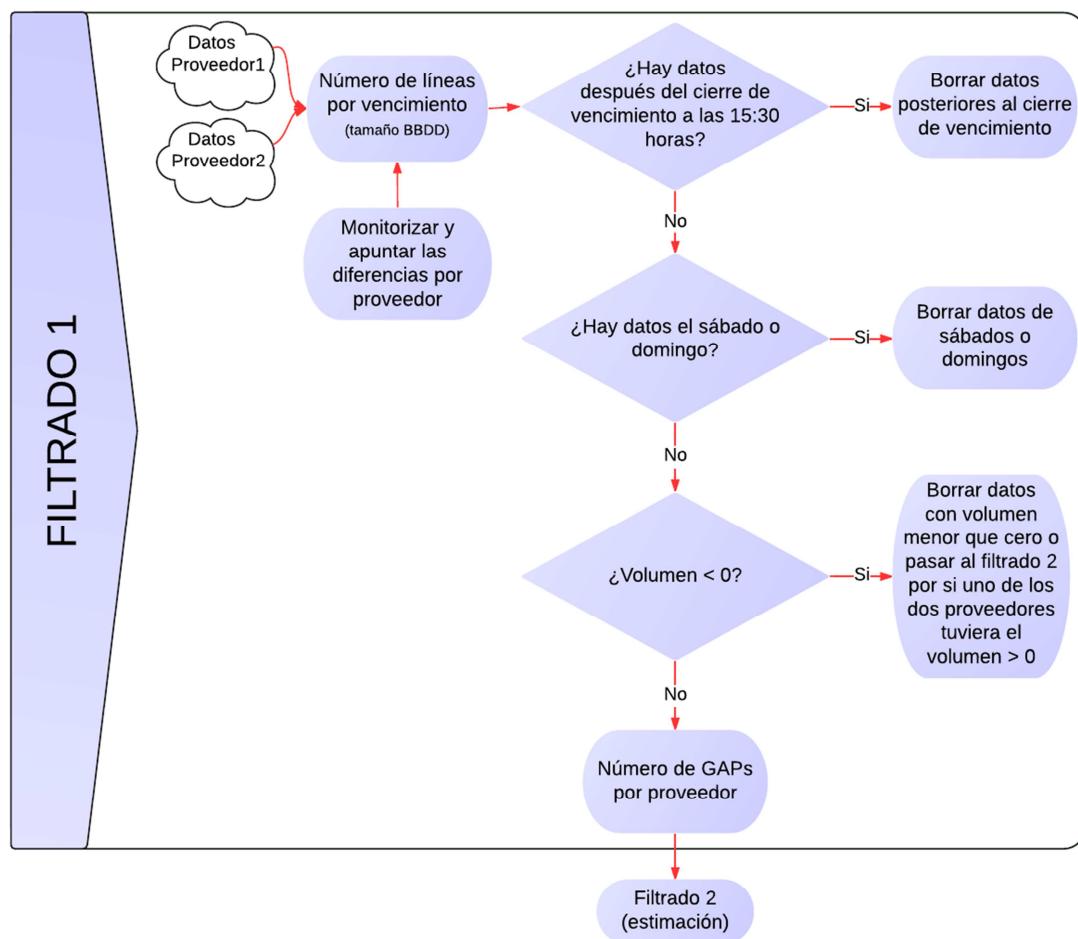


Figura 27: Filtrado 1 – preprocesado de datos - Proceso ATurel

- **Filtrado 2 (estimación)**

Nos referimos en el diagrama a proveedor1 como p1 y a proveedor2 como p2. El ajuste por mínimos cuadrados, tal como hemos explicado, será por medio de una curva polinómica lo más ajustada posible (sobreoptimizada al tratar temas de estimación y no de predicción). Por tanto, aplicaremos una estimación con un polinomio de grado nueve. Además, para poder dividir los resultados en tendencia bajista | alcista | lateral procederemos con el cálculo de la recta tangente por cada punto de la curva de ajuste. La pendiente de la recta tangente nos marcará estos resultados pero como no vamos a obtener exactamente pendiente cero en un completo tramo lateral, ésta se debería optimizar. Como no atañe a este trabajo, los problemas ya mencionados sobre justificar todas las decisiones durante el desarrollo del simulador podría abarcar un Proyecto Final de Carrera o incluso Tesis Doctoral por cada decisión tomada a modo de suposición en algunas fases del presente trabajo, los tramos laterales serán capturados en los máximos y mínimos del gráfico.

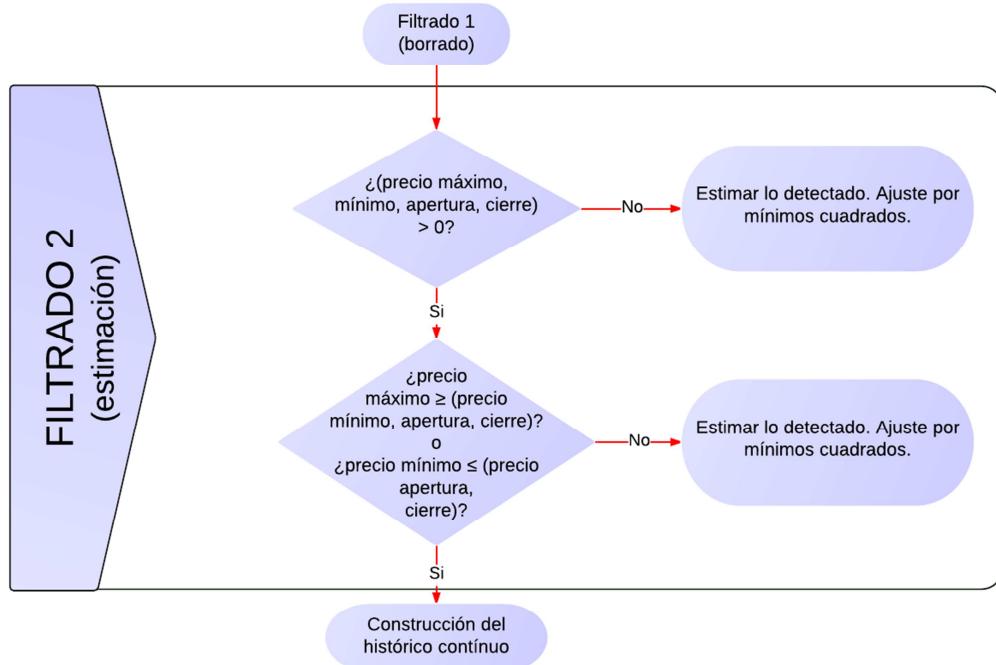


Figura 28: Filtrado 2 – preprocessado de datos - Proceso ATurel

- **Construcción del histórico continuo**

Después de preprocessar los datos de los vencimientos, reflejado en el ANEXO II, el período de máxima negociación considerado es de las 17:00 horas hasta las 22:00 horas. Al final el histórico continuo obtenido será utilizado una parte como muestra IS (2010-2012) y otra como muestra OOS (2013).

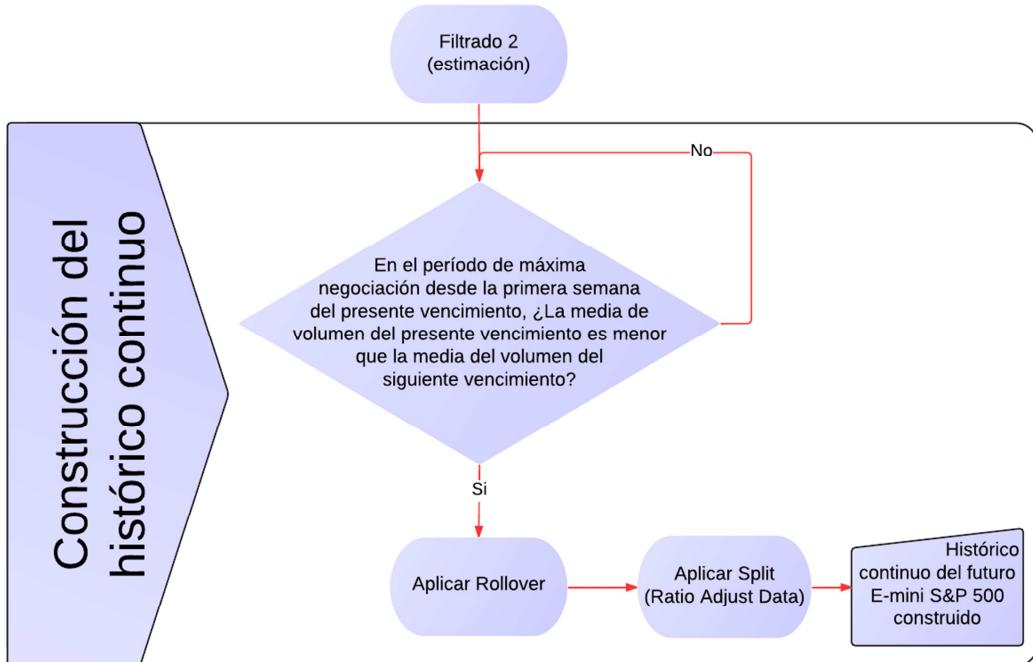


Figura 29: construcción del histórico continuo – Preprocesado de datos - Proceso ATurel

Continuamos con la fase 2,

'Optimización Sistema Algorítmico de Trading' + 'filtros':

El diagrama muestra el caso típico de estudio con ejecución secuencial (bucles pero programado en paralelo para aprovechar todos los *cores*, núcleos). En MATLAB nos ayudarán las sentencias *parfor* y *matlabpool* para paralelizar la ejecución en todos los *cores* habilitados. Se mostrarán los resultados acumulados, los cuales utilizaremos para continuar con el siguiente elemento del Proceso ATurel pero también dividiremos los resultados según en tendencia alcista | bajista | lateral con el fin de comprobar dónde se comporta mejor el sistema optimizado. Aplicaremos un muestreo con *timeframe* = 15 minutos en base al histórico continuo de *timeframe* = 1 minuto obtenido en la fase 1. Es decir, aunque durante la optimización podríamos ajustar el *timeframe* como si de un parámetro más se tratase, no lo vamos a hacer porque los resultados pueden sufrir distorsión debido a que la muestra IS varía. Por otro lado se suele considerar que si el estudio de un Sistema Algorítmico de *Trading* para un *timeframe* fijo obtiene buenos resultados entonces para demostrar robustez se debería tener resultados proporcionales en otros *timeframes*.



Figura 30: optimización Sistema Algorítmico de Trading + Filtros - Proceso ATurel

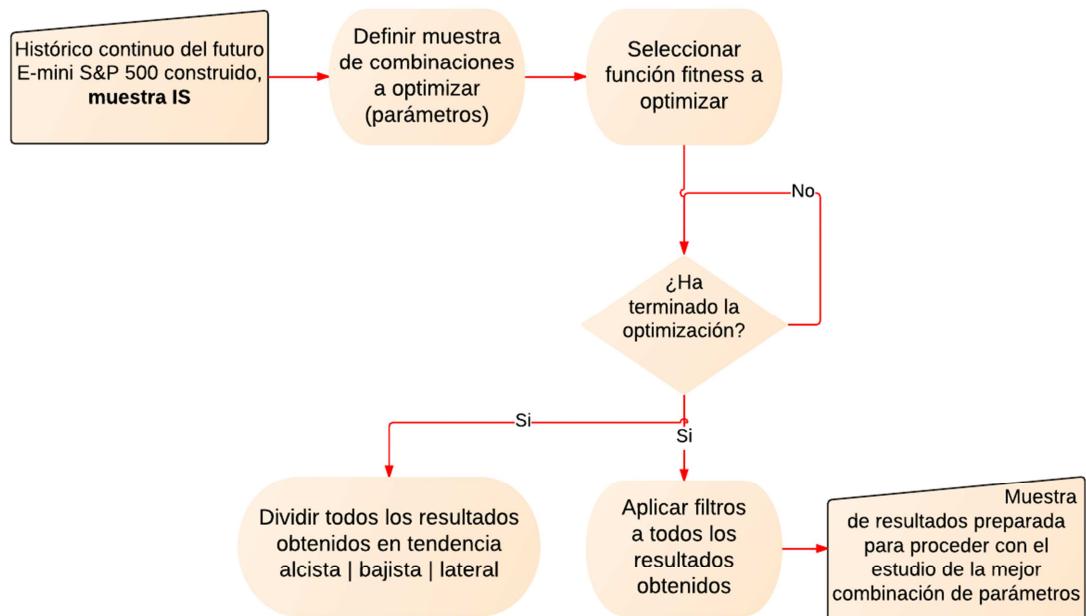


Figura 31: optimización Sistema Algorítmico de Trading + Filtros, en detalle - Proceso ATurel

'PBO'

Procedimiento ya explicado en el punto 2.6.2. Veremos en el modo de evaluación que utilizaremos los rendimientos de la función *fitness* en la muestra OOS en base a los resultados obtenidos durante optimización IS. Comprobaremos que a medida que aumenta la probabilidad de sobreoptimización (overfitting), el rendimiento IS es mucho mayor que OOS, es decir, que éste disminuye considerablemente.



Figura 32: PBO - Proceso ATurel

'Mejor combinación de parámetros'

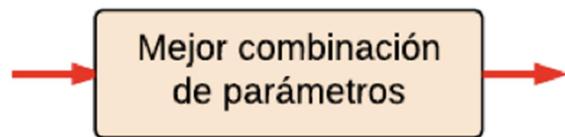


Figura 33: mejor combinación de parámetros - Proceso ATurel

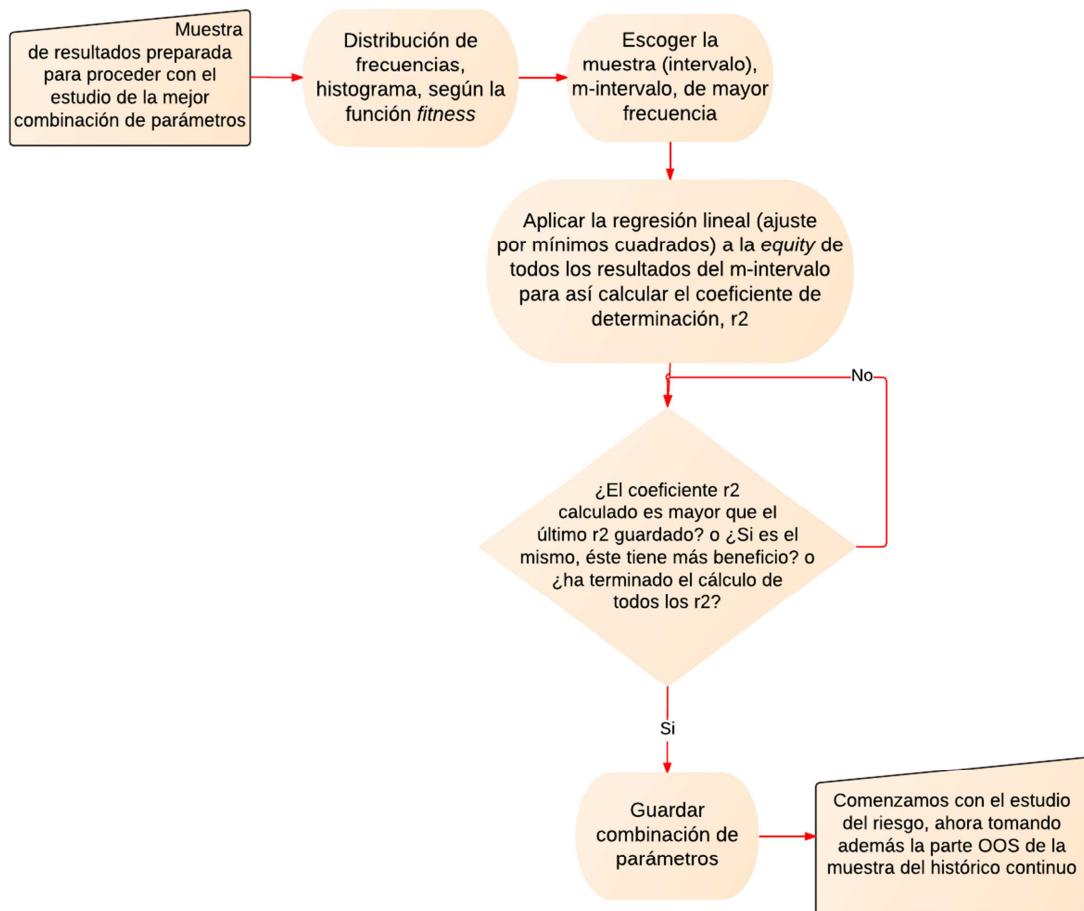


Figura 34: mejor combinación de parámetros en detalle - Proceso ATurel

Finalmente la fase 3,

'Monte Carlo'

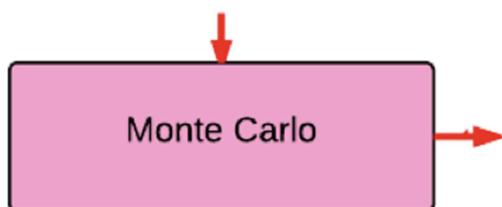


Figura 35: Monte Carlo - Proceso ATurel

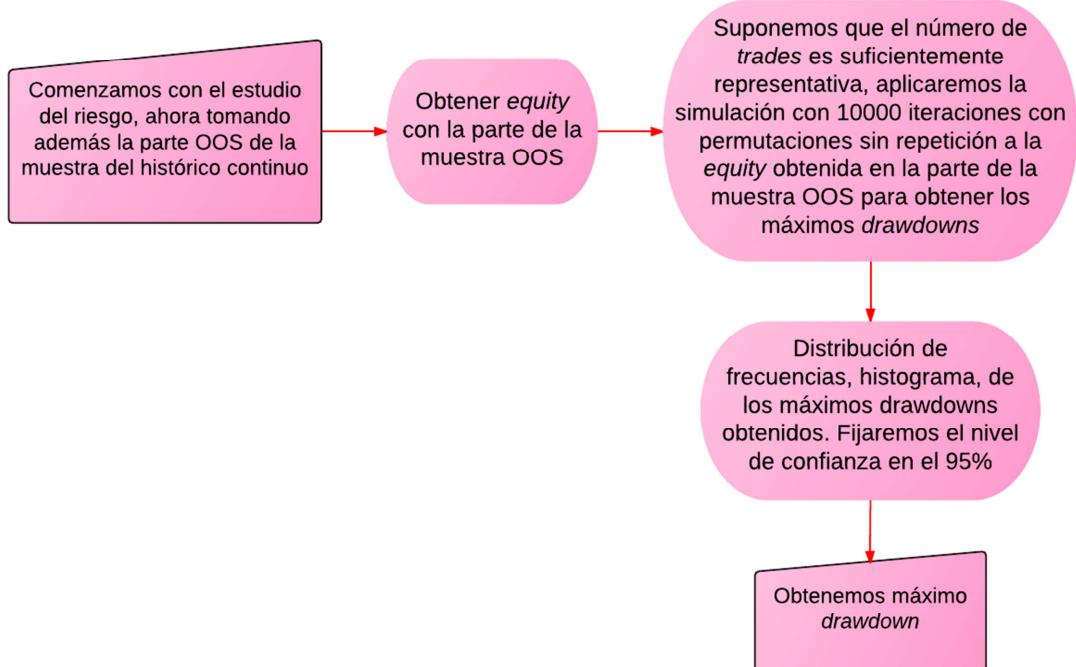


Figura 36: Monte Carlo detallado - Proceso ATurel

'Ratio Calmar'

No tiene más misterio que realizar un cálculo matemático. La *equity* obtenida de la parte de la muestra OOS nos muestra, entre otros, el supuesto beneficio obtenido y con el cálculo del máximo *drawdown* gracias a la técnica del método Monte Carlo calcularemos el Ratio Calmar y así veremos el mejor rendimiento beneficio/riesgo.

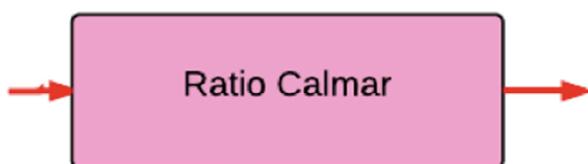


Figura 37: Ratio Calmar - Proceso ATurel

3.1.1. Aspectos metodológicos

Después de haber visto la metodología seguida en el punto anterior y el detalle de sus elementos funcionales, pasamos entonces a explicar cada uno de los elementos no funcionales (cajas no rectangulares) que forman el proceso ATurel;

'Datos mercados financieros (históricos)'



Figura 38: datos mercados financieros - Proceso ATurel

Como ya hemos indicado, en el presente trabajo estudiaremos un preprocesado de datos con dos proveedores (proveedor1 y proveedor2). Para cada proveedor, trabajaremos con los vencimientos del futuro E-mini S&P 500 desde el 4 de enero de 2010 hasta el 20 de diciembre de 2013 en los siguientes rangos:

- Vencimiento marzo: desde el 1 de diciembre del año n hasta el tercer viernes de marzo del año n+1.
- Vencimiento junio: desde el 1 de marzo del año n hasta el tercer viernes de junio del año n.
- Vencimiento septiembre: desde el 1 de junio del año n hasta el tercer viernes de septiembre del año n.
- Vencimiento diciembre: desde el 1 de septiembre del año n hasta el tercer viernes de diciembre del año n.

'Resultados Fase 1'



Figura 39: resultados Fase 1 - Proceso ATurel

Entre otros resultados que estarán disponibles en el ANEXO II tanto para comprobar la demostración de las tres partes del preprocesado de datos como las diferencias entre los proveedores, el resultado más importante será el histórico continuo del futuro E-mini S&P 500 que se ha construido según la metodología explicada en el punto anterior.

'Resultados Fase 2'



Figura 40: resultados Fase 2 - Proceso ATurel

Entre otros resultados que estarán disponibles en el ANEXO III para comprobar la demostración de los pasos seguidos, los resultados más importantes serán; 1. La probabilidad de sobreoptimización, 2. La mejor combinación de parámetros y 3. Medidas de rendimiento computacional para así compararla más adelante con el experimento veintisiete.

'Resultados Fase 3'



Figura 41: resultados Fase 3 - Proceso ATurel

Entre otros resultados que estarán disponibles en el ANEXO IV, el resultado más importante será aquel que verifique cuál es el mejor experimento, en función del número de combinaciones y demás elementos a testear, en función del beneficio/riesgo. Llegados a este punto las conclusiones estarán prácticamente resueltas.

3.1.1.1. Diseño de experimentos realizados

Voy a emplear dieciocho experimentos más uno sobre el problema de la sobreoptimización para dar las conclusiones necesarias para demostrar los objetivos redactados en el punto 1.6. Como se puede ver a continuación, se dividen los experimentos de la siguiente forma:

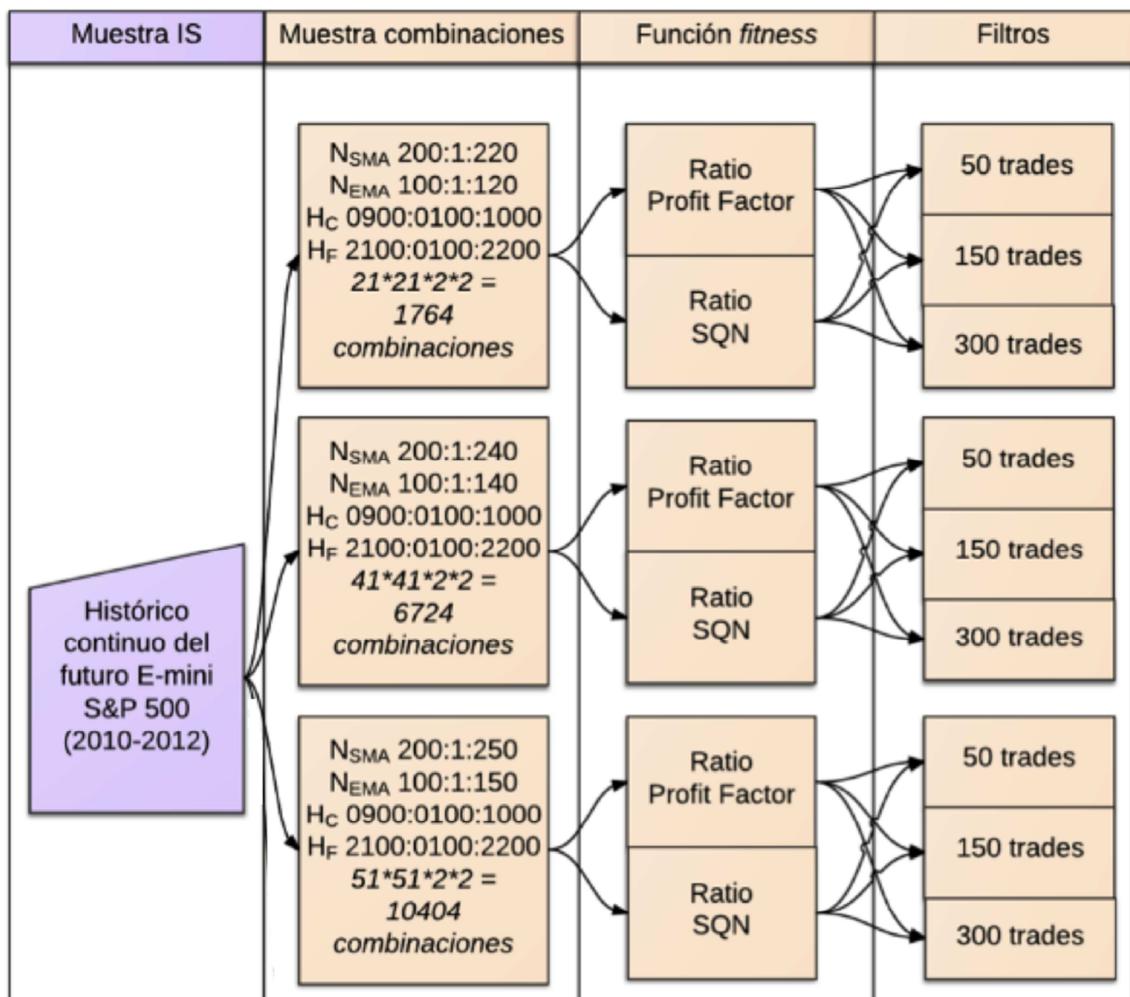


Figura 42: desglose de los dieciocho experimentos

No indico los pasos siguientes del Proceso ATurel porque no genera más combinaciones para demostrar de dónde vienen los $3*2*3 = 18$ experimentos. Obviamente el resto de fases del Proceso ATurel se elaboran a continuación de lo mostrado en la anterior figura según la metodología y los aspectos metodológicos ya explicados. Expreso los valores de los parámetros de la forma `valorInicio:salto:valorFinal`, las horas están expresadas en formato 24 horas y el `timeframe` empleado en el histórico continuo de quince minutos.

Para demostrar el objetivo 1 tomaré un rango de parámetros fijo, en este caso aquel que contenga el mayor número de combinaciones (10404). A continuación muestreará de dos en dos teniendo entonces 5202 combinaciones y finalmente volveré a muestrear de dos en dos respecto a la anterior matriz o de cuatro en cuatro respecto a la primera matriz (10404 combinaciones) para tener así 2601 combinaciones. obtendré la probabilidad de sobreoptimización y los rendimientos IS y OOS.

3.1.1.2. Evaluación

La evaluación de la fase 1 queda explicada de forma clara en el diagrama de flujos de la página 57. La principal evaluación para el resto de resultados de la fase 2 y 3 será comprobar la diferencia de los rendimientos IS y los rendimientos OOS. Es decir, los gráficos y ratios obtenidos en la muestra IS en comparación con los obtenidos en la muestra OOS será la demostración objetiva de si el resultado es bueno o es malo. Si el rendimiento de la muestra IS es muy buena pero el rendimiento de la muestra OOS es muy mala (por ejemplo porque se obtienen pérdidas) nos estará alertando que la probabilidad de sobreoptimización aumenta. A razón del objetivo 2 bis explicado en el punto 1.6, también podremos evaluar si el método Monte Carlo es correcto o no comparando el máximo *drawdown* de la muestra OOS antes y después de aplicar dicho método. Se supone que el máximo *drawdown* debe ser mayor después del método Monte Carlo.

Bien cabe destacar, lógicamente, que las conclusiones finales serán justificadas en orden inverso a la realización del Proceso ATurel. Es decir, hasta que no tengamos los resultados de la fase 3, no podemos verificar por completo si el experimento es bueno o malo. Aunque la probabilidad de sobreoptimización será un resultado intermedio para evaluar posibles alertas en los resultados de la fase 3.

3.1.2. Experimentos extras elaborados

Tenemos tres experimentos extras para demostrar por un lado la importancia de la fiabilidad de los datos y por otro lado si merece la pena, computacionalmente hablando, implementar paralelización para optimizar en función del rendimiento del Sistema Algorítmico de Trading. A continuación detallamos los experimentos a desarrollar:

- **Experimento veinte:** Despues de escoger el mejor experimento (beneficio/riesgo) de los veinticuatro posibles explicados en el punto 3.1.1.1, éste lo volveremos a realizar pero esta vez sin un histórico continuo elaborado. Es decir, me salto la fase 1 del Proceso ATurel y simplemente realizo la unión de los vencimientos (rollover con fecha fija, explicado en el punto 2.5.) y sin aplicar tampoco un ajuste (*split*). Por tanto, desarrollaré la fase 2 y la fase 3 del Proceso ATurel en base a un histórico continuo del proveedor1. Veremos que los rendimientos IS y OOS varían.
- **Experimento veintiuno:** Lo mismo que lo explicado en el anterior experimento pero utilizando como fuente de datos el proveedor2.
- **Experimento veintidós:** Como trabajo extra para lidiar con los rendimientos computacionales ejecutaré la mayor combinación de parámetros de los anteriores veinticuatro experimentos de forma no en paralelo y también sin reservar memoria como aconsejábamos para comprobar diferencias computacionales.

Podríamos suponer en un principio que los rendimientos de los últimos seis experimentos a razón de un millón de combinaciones no van a resultar ninguno como mejor resultado, lo veremos en los resultados cuando compararemos los rendimientos IS y OOS.

3.2. Material

El material más importante utilizado es la aplicación de conocimientos durante la experiencia de los últimos años de formación en la Escuela Técnica Superior de Ingenieros de Telecomunicación. Lógicamente para desarrollar el presente trabajo he tenido que utilizar tanto programas informáticos como dispositivos hardware para ejecutar software.

3.2.1. Software

La mayoría del trabajo está implementado en el entorno de desarrollo matemático MATLAB. Se trata de una plataforma de cálculo muy potente por la cual podemos utilizar herramientas de estadística gracias a Statistics toolbox, herramientas financieras gracias a Financial toolbox y herramientas de optimización gracias a Optimization toolbox.

El software restante empleado ha sido desarrollado en el lenguaje de programación C, en concreto en la parte del preprocesado de los datos para adaptar formatos de lectura de ficheros para MATLAB.

No podemos olvidar a los dos proveedores de datos que ofrecen el agente principal para poder testear el software necesario y así concluir justificando los objetivos según los resultados obtenidos.

3.2.2. Hardware

A estas alturas repetir que es crucial disponer de máquinas potentes es innecesario por el nivel de concienciación transmitido al lector. Las características técnicas de las máquinas empleadas durante el trabajo son:

Por un lado desarrollo con mi máquina inseparable desde que entré en la E.T.S.I. Telecomunicación y que adquirí por cortesía de mis padres:

El primer modelo Intel Core 2 Duo de 2.2 GHz de Macintosh y un aumento de RAM hasta los 4 GB pero las características del hardware del modelo aguanta hasta 3 GB.

Por otro lado para las simulaciones utilizo una máquina más potente que he podido realizar gracias al equipamiento que cuenta el Grupo de Aplicaciones de Procesado de Señal (GAPS) del Departamento de Señales, Sistemas y Radiocomunicaciones (SSR) de la E.T.S.I Telecomunicación – Universidad Politécnica de Madrid.

Un breve resumen de sus características técnicas*:

- Procesador
 - 6 núcleos (cores) físicos Intel i7 Core 2 Duo de 3.2 GHz cada uno. Es decir, 12 cores en total.
 - 64 bits
- Kernel: Linux 3.16.0-23-generic
- Distribución Ubuntu 14.10
- RAM: 24089 MB
- Memoria compartida (swap): 24552 MB
- Cache: 12.288 MB

*Información de CPU (cat /proc/cpuinfo), información de memoria (cat /proc/meminfo)

4. Resultados

Recomiendo encarecidamente al lector revisar siempre que pueda el informe detallado en los ANEXOS II, III y IV para poder seguir con mayor detalle todos los resultados obtenidos.

Según como habíamos explicado en el punto anterior, los resultados de la fase 1 vendrán determinados principalmente por la construcción y resultado del histórico continuo ya filtrado, con el *rollover* y *split* aplicado.

En cuanto a la fase 2 y 3, muestro los mejores experimentos en función del procedimiento explicado en el punto anterior; Muestra de combinaciones, función *fitness* empleada y filtro por número de contratos. En concreto a los tres experimentos más relevantes, uno por cada combinación de parámetros analizado, nos ayudarán a obtener las conclusiones más robustas de todo el trabajo realizado junto con el experimento del objetivo 1 sobre la probabilidad de sobreoptimización. El procedimiento de presentación de resultados será; nombre experimento, muestra de combinaciones seleccionada, función *fitness* aplicada, filtro utilizado, tiempo de ejecución de la optimización, carga media del sistema y consumo RAM, PBO obtenido, nube de puntos del valor de la función *fitness* IS (muestra de los mejores 2000 resultados), nube de puntos del valor de la función *fitness* OOS, histograma de los valores de la función *fitness* IS, mejor combinación de parámetros (mejor *equity* IS), retorno - máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS, número de *trades* en tramos alcistas | bajistas | laterales, gráfica de la *equity* IS superpuesta con la *equity* OOS de la mejor combinación de parámetros, máximo *drawdown* (IS OOS y por Monte Carlo) y *Ratio Calmar*.

Los resultados de los experimentos extras serán expuestos de la misma forma que los ya mencionados en la fase 2 y 3 pero lógicamente sin indicar identificación que necesitara del experimento uno al dieciocho. Sin olvidar que también presentaré las gráficas de los gráficos continuos sin realizar el preprocesado de los datos de la fase 1.

4.1. Fase 1

El filtrado 1 destaca por descubrir ciertas irregularidades en el proveedor 2, he detectado errores del tipo; Datos algunos domingos, no cierre del vencimiento a las 15:30 horas y tres vencimientos (ESM13, ESU12, ESU13) con un dato a las 15:31 horas y con un volumen igual a uno.

En el filtrado dos no he encontrado irregularidades, si acaso los datos de volumen son distintos en los dos proveedores. Este dato, aun no siendo tan preciso puesto que lo trata a modo de paquete (no sabemos los *trades* de compra o de venta por ejemplo), es bastante importante para comprender cómo se mueve la oferta y la demanda pero no me da la suficiente confianza después de comprobar que son distintos para los dos proveedores.

En cuanto a la construcción del histórico continuo, La tendencia para realizar el *rollover* del futuro E-mini S&P 500 es justamente una semana antes al día marcado oficialmente, siendo éste el tercer viernes del mes.

Por último los tramos laterales que tomaremos en cuenta son (insisto que el procedimiento seguido es mejorable):

Tramo	Inicio (yyyyMMdd HHmm)	Fin (yyyyMMdd HHmm)
1	20100104 1900	20100121 0726
2	20100412 2134	20100430 0633
3	20100715 1420	20100818 0308
4	20110228 0403	20110309 2006
5	20110331 0317	20110418 1717
6	20110815 2014	20110927 1127
7	20111031 0329	20111115 1442
8	20120730 0608	20120802 1800
9	20121024 0759	20121106 0854
10	20121224 0759	20121204 1503

Tabla 3: tramos laterales, histórico continuo filtrado y construido

Pasemos a ver unas representaciones gráficas de cómo ha quedado el histórico continuo filtrado con el *rollover* y el *split* aplicado. También la muestra IS y la muestra OOS por separado.

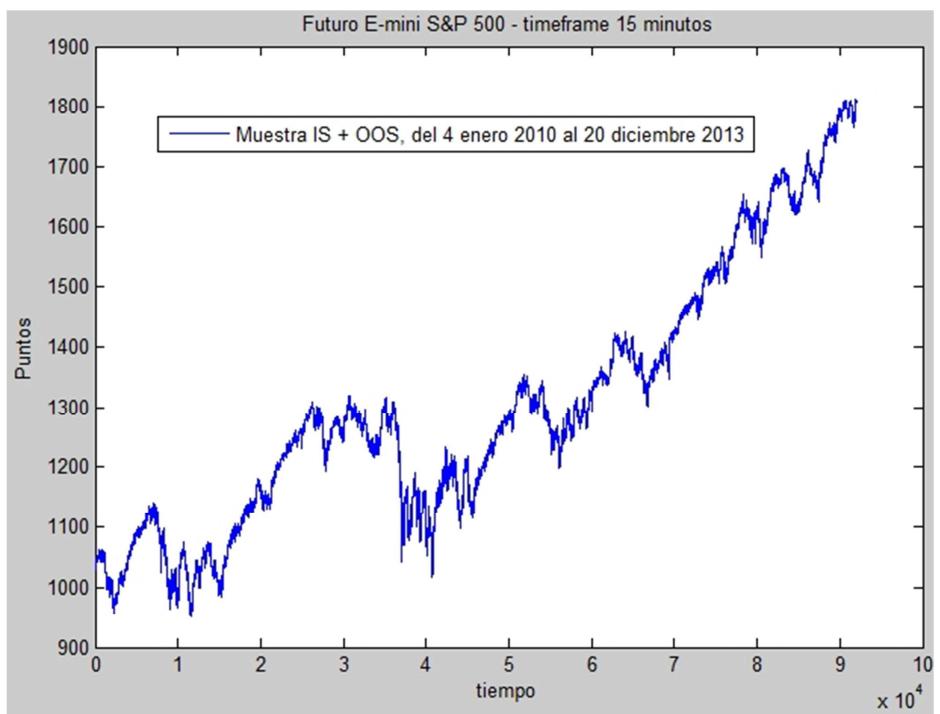


Figura 43: muestra IS+OOS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

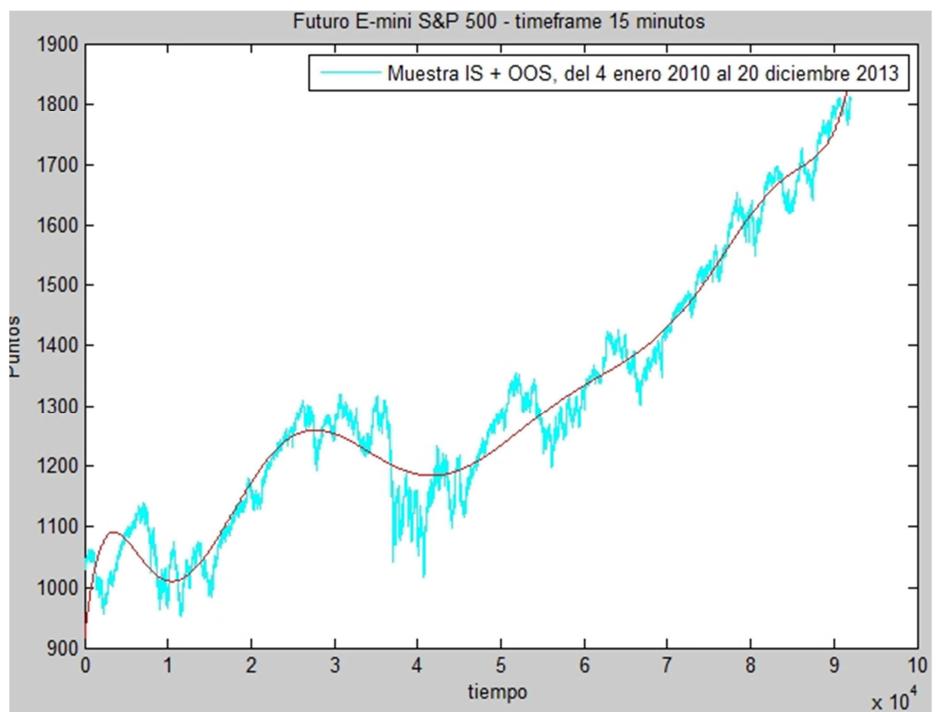


Figura 44: muestra IS + OOS, ajuste – precio cierre – E-mini S&P 500 - 15 min

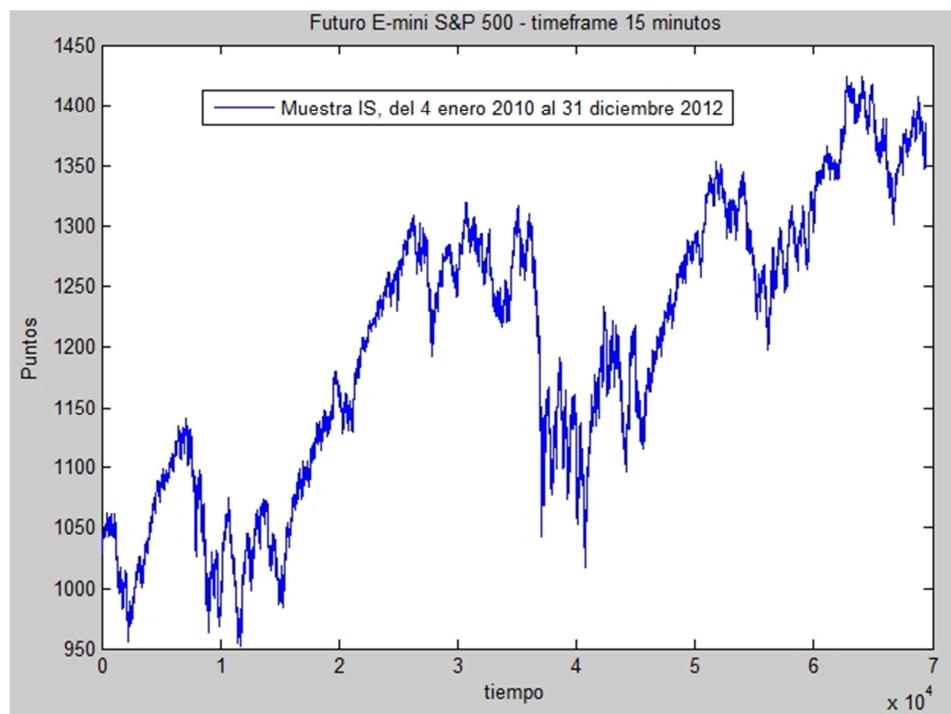


Figura 45: muestra IS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

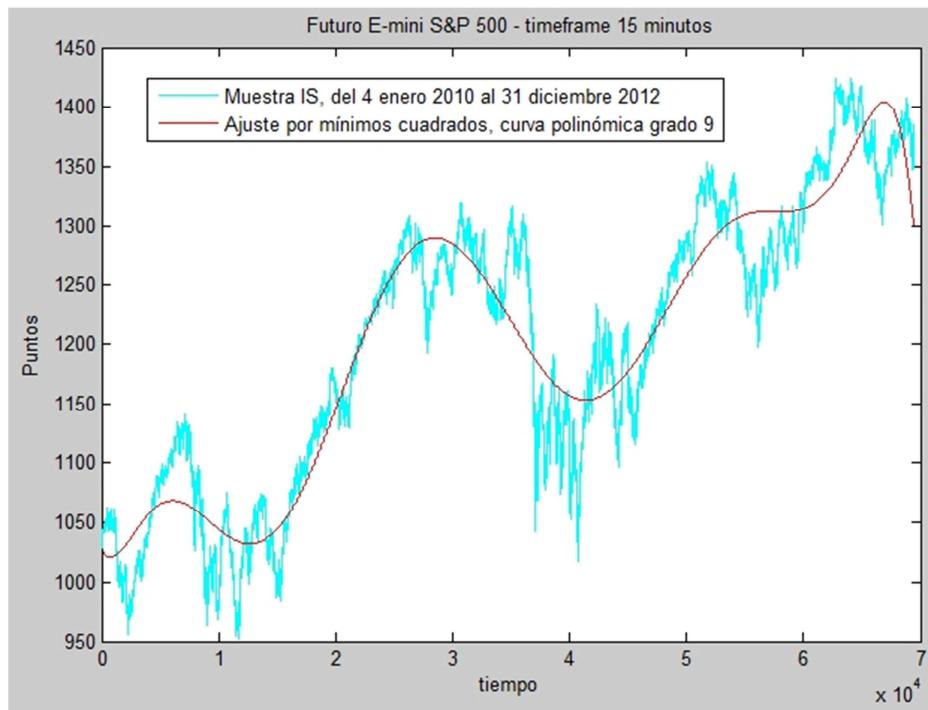


Figura 46: muestra IS, ajuste – precio cierre – E-mini S&P 500 - 15 min

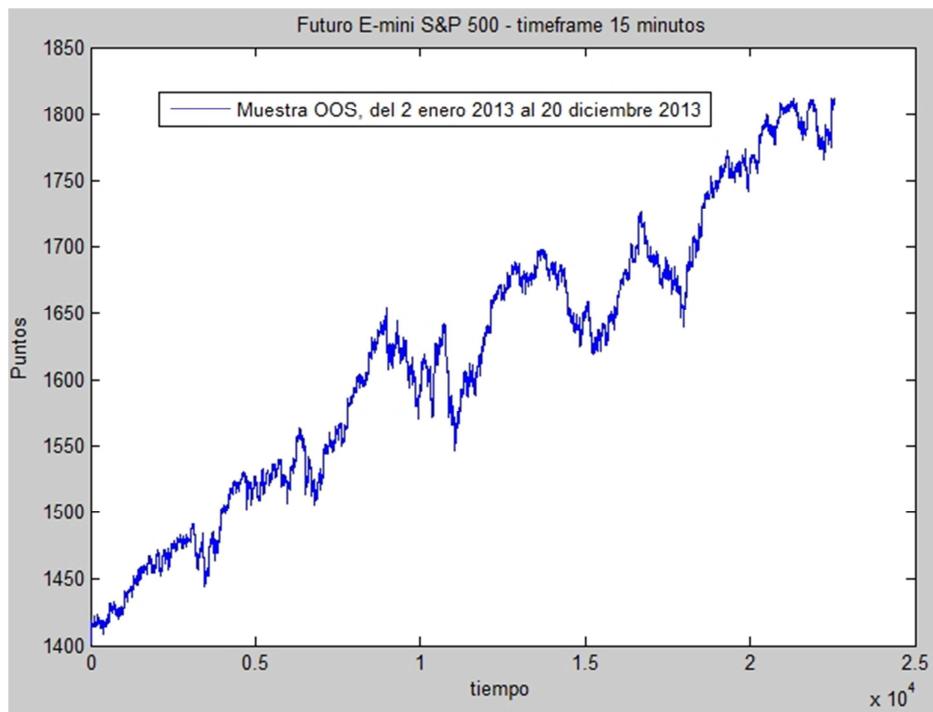


Figura 47: muestra OOS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

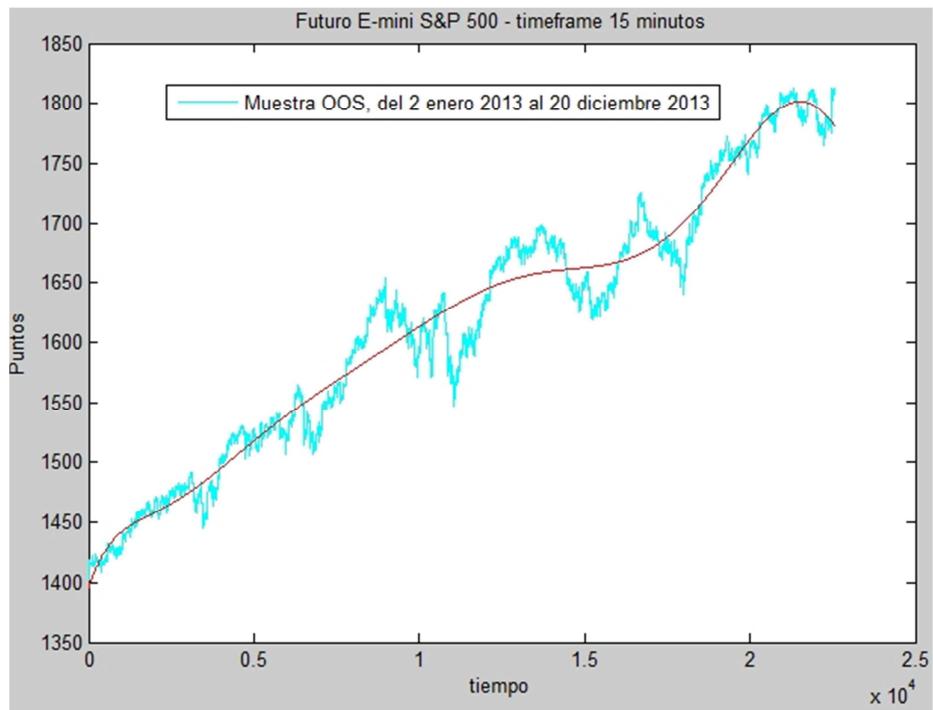


Figura 48: muestra OOS, ajuste – precio cierre – E-mini S&P 500 - 15 min

4.2. Fase 2

Tomamos en cuenta que el histórico continuo filtrado y muestreado en 15 minutos ocupa en un vector fila de 69432 columnas.

Un detalle importante referente a lo que mencionábamos en el punto 2.5 sobre el procedimiento Map-Reduce. Para el estudio de los experimentos había considerado guardar en una matriz;

[vector P&L][número de *trades*][valor función fitness SQN] [valor función fitness SR]
[valor de los 4 parámetros][máx *drawdown*][máximo beneficio/pérdida]

Esto sería para el experimento 13, 14, 15, 16, 17 y 18: $10404 * (69432 + 9) * 8 = 5779713312$ B = 5.78 GB de memoria RAM a reservar como mínimo.

Pues bien, la diferencia de calcularlo TODO de golpe para llenar la matriz y el cálculo de toda la matriz menos la columna de máximo *drawdown* aumenta en un 10483% el tiempo de ejecución. La operación para el cálculo del máximo *drawdown* es crítica, como dicho dato lo necesitaré llegados a la fase 3 y como encima en la fase 3 analizamos el máximo *drawdown* de una sola *equity* a razón de la mejor combinación de parámetros ¿Para qué calcular según el proceso seguido el máximo *drawdown* de las 10404 combinaciones de parámetros para los experimentos citados? Como moraleja aconsejo depurar el programa a ejecutar para saber consumo de CPU, las llamadas al sistema y procesos padres e hijos entre otros para así optimizar el consumo, porque el tiempo es oro. En MATLAB nos puede ayudar la sentencia *profile*.

Otro ejemplo muy sencillo sobre Map-Reduce podría ser:

¿El precio mínimo \leq precio máximo, precio cierre y precio apertura?

1. Sin utilizar Map-Reduce: Comprobar que la condición es cierta por cada precio analizado.
2. Utilizando Map-Reduce
 - a. Map: Restar precio mínimo con cada uno de los otros precios, construir un vector según el precio.
 - i. Vector 1: Precio mínimo – precio máximo
 - ii. Vector 2: Precio mínimo – precio apertura
 - iii. Vector 3: Precio mínimo – precio cierre
 - b. Reduce: Si algún dato de cada vector es > 0 entonces no se ha cumplido.

El tiempo de ejecución en este tipo de operaciones para grandes bases de datos se nota, y bueno esto sería uno de los ejemplos que bases de datos no relacionales como MongoDB realizan para disminuir el tiempo en las lecturas. Objetivo: realizar todas las operaciones matemáticas intermedias posibles para que la operación final (objetivo final) sea solamente una operación simple (una suma por ejemplo) o una condición simple. Último ejemplo sobre una búsqueda con MongoDB con procedimiento Map-Reduce.

Objetivo: Buscar cuántas palabras ‘casa’ hay en un texto de varios gigabytes.

Map: Construir vector para añadir un uno cada vez que se encuentra la palabra ‘casa’.
Reduce: Sumar número de unos.

Con este ejemplo tan tonto se pueden conseguir rendimientos computacionales sustanciales y por eso mi gran estima a los matemáticos que citaba en el punto 1.4 cuando comentaba que ellos pueden encontrar las operaciones más simples por pasos para obtener el resultado final lo más rápido posible.

Es importante tomar en cuenta que no hay diferencias, entre los experimentos, a razón el filtrado mínimo por número de contratos debido a que ha dado la casualidad que el presente sistema en la muestra de datos (IS) ejecutada realiza más de 300 operaciones (*trades*) siempre.

Para tener una referencia computacional en los resultados obtenidos, mostramos el mapa de la RAM del sistema nada más arrancar:



Figura 49: consumo de *RAM* inicial – Sistema

Y una vez arrancada la aplicación MATLAB:



Figura 50: consumo de *RAM* inicial con MATLAB arrancado – Sistema

Sin más dilación comenzamos mostrando los mejores experimentos en función de la mejor combinación de parámetros:

Nombre del experimento: Experimento 1,2 y 3

Número de combinaciones: 1764

Función *fitness*: SQN

Filtro: 50, 100 y 300

Tiempo de ejecución de la optimización (paralelo) = 2.76 minutos

Tiempo de ejecución de la optimización (no paralelo) = 3.09 minutos



Figura 51: consumo *RAM* – Experimento 1, 2 y 3

Como mínimo necesitamos una variable que ocupa $69441 * 1764 * 8 = 979.95$ MB

1	[1.9%]	4	[0.0%]	7	[100.0%]	10	[100.0%]
2	[100.0%]	5	[100.0%]	8	[0.5%]	11	[0.0%]
3	[1.4%]	6	[3.7%]	9	[100.0%]	12	[100.0%]

Figura 52: carga del sistema – 6 cores – Experimento 1, 2 y 3

Load average: 2.51 3.58 4.01

Figura 53: carga media del sistema – Experimento 1, 2 y 3

PBO = 100/252 (39.68 %)

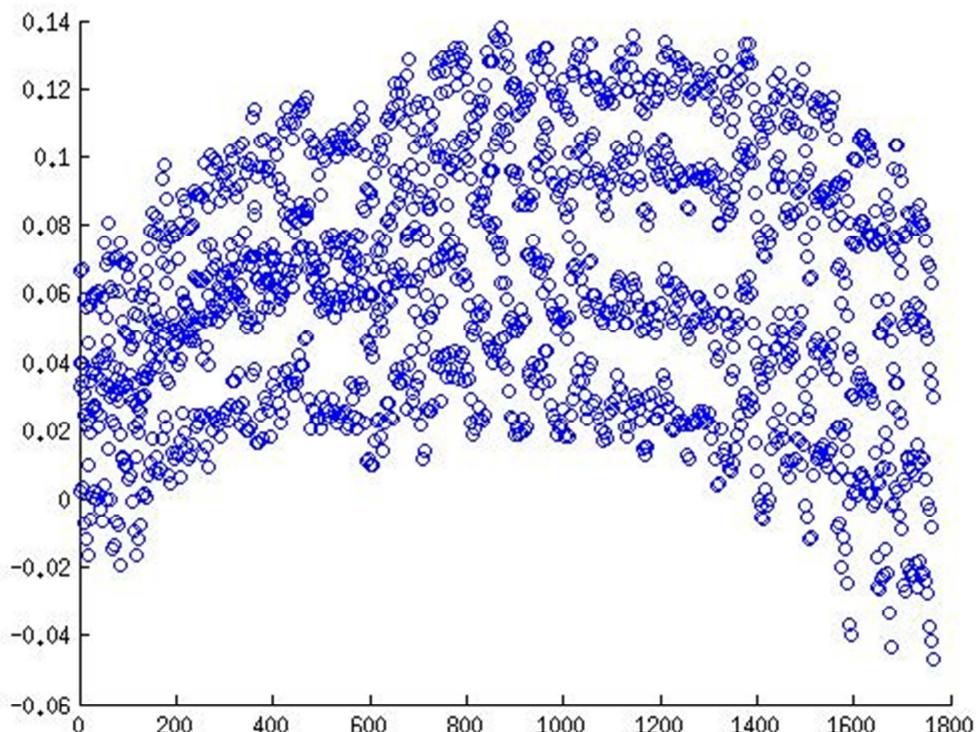


Figura 54: nube de puntos del resultado de la función *fitness* IS – Experimento 1, 2 y 3

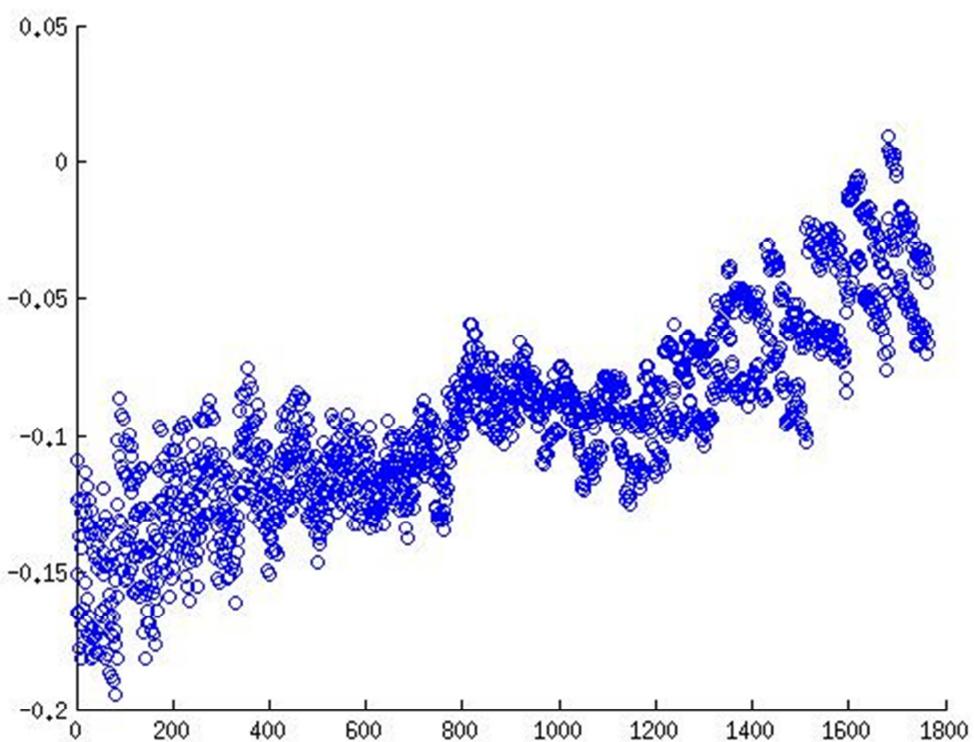


Figura 55: nube de puntos del resultado de la función *fitness* OOS – Experimento 1, 2 y 3

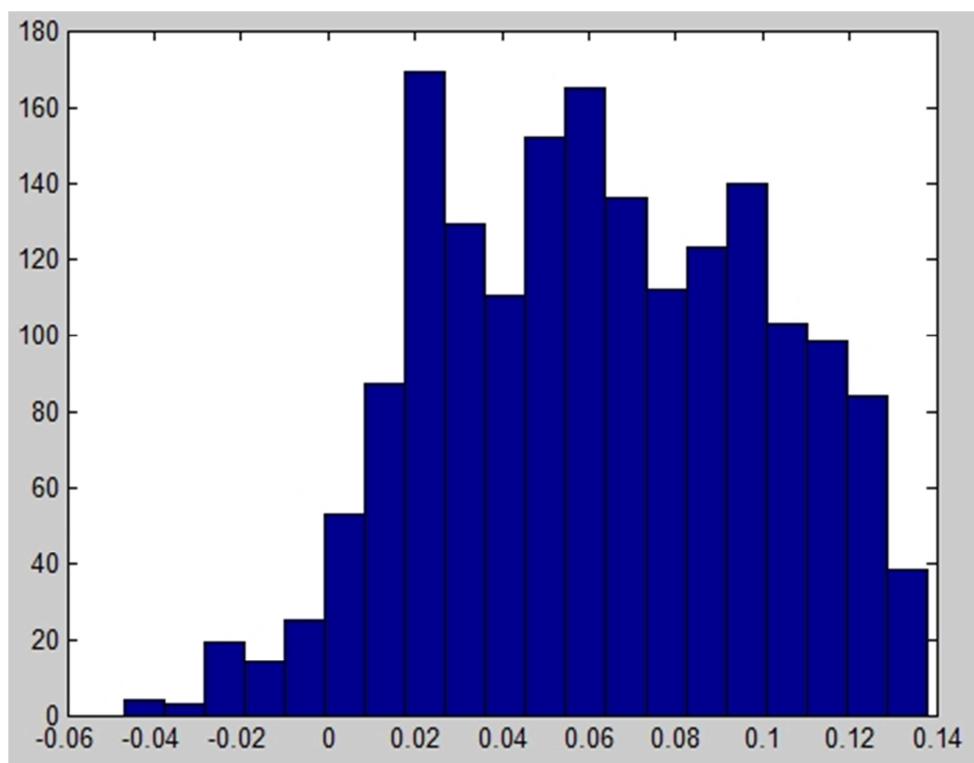


Figura 56: histograma de los valores de la función *fitness* IS – Experimento 1, 2 y 3

periSMA	periEMA	hComienzo	hFin	Coeficiente de Determinación
200	102	0900	2100	0.3183

Tabla 4: mejor combinación de parámetros – Experimento 1, 2 y 3

Retorno IS	Máximo drawdown IS	Función fitness IS	Retorno OOS	Máximo drawdown OOS	Función fitness OOS
3202.7	1975	0.0245	-11262	12610	-0.1812

Tabla 5: retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 1, 2 y 3

Número total de trades	Trades en tramo alcista	Trades en tramo bajista	Trades en tramo lateral
969	568	240	161

Tabla 6: *trades* en tramo alcista | bajista | lateral – Experimento 1, 2 y 3

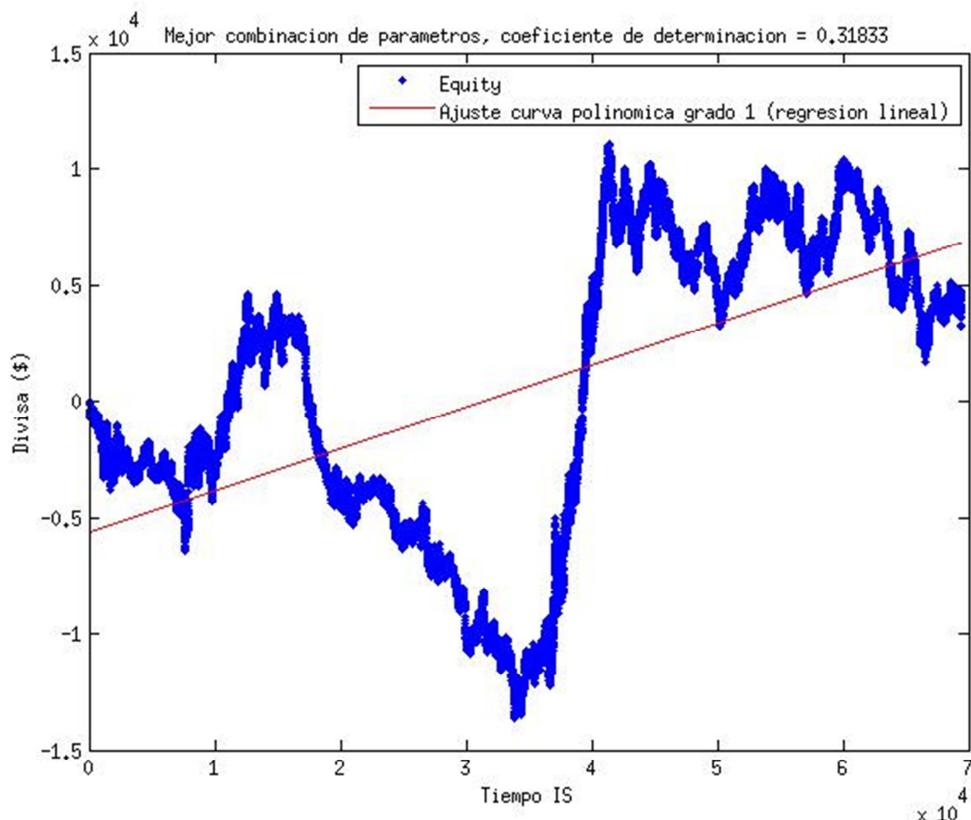


Figura 57: equity IS – Experimento 1,2 y 3

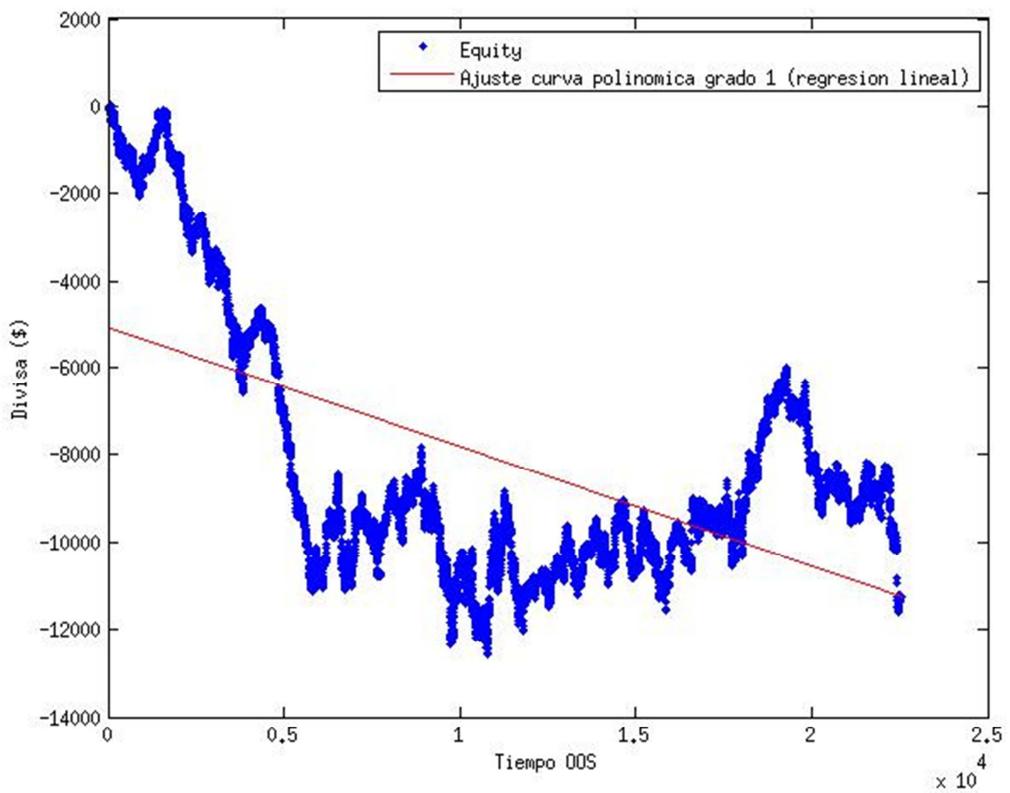


Figura 58: *equity* OOS – Experimento 1,2 y 3

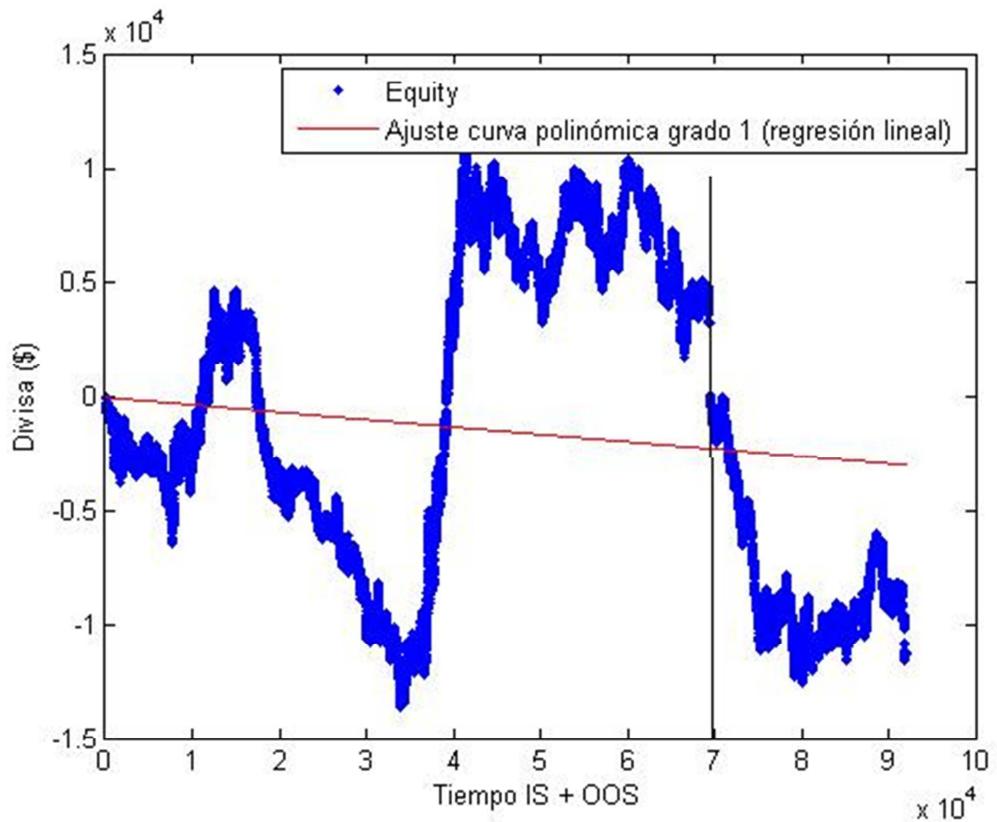


Figura 59: *equity* IS + OOS – Experimento 1,2 y 3

Nombre del experimento: Experimento 10, 11 y 12

Número de combinaciones: 6724

Función *fitness*: SR

Filtro: 50, 150 y 300

Tiempo de ejecución de la optimización (paralelo) = 11.86 minutos

Tiempo de ejecución de la optimización (no paralelo) = 12.31 minutos



Figura 60: consumo RAM – Experimento 10, 11 y 12

Como mínimo necesitamos una variable que ocupa $69441 * 6724 * 8 = 3.74$ GB

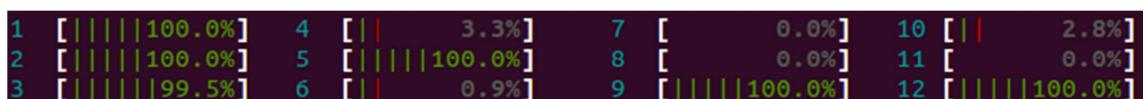


Figura 61: carga del sistema – 6 cores – Experimento 10, 11 y 12

Load average: 6.04 5.39 4.34

Figura 62: carga media del sistema – Experimento 10, 11 y 12

PBO = 117/252 (46.43%)

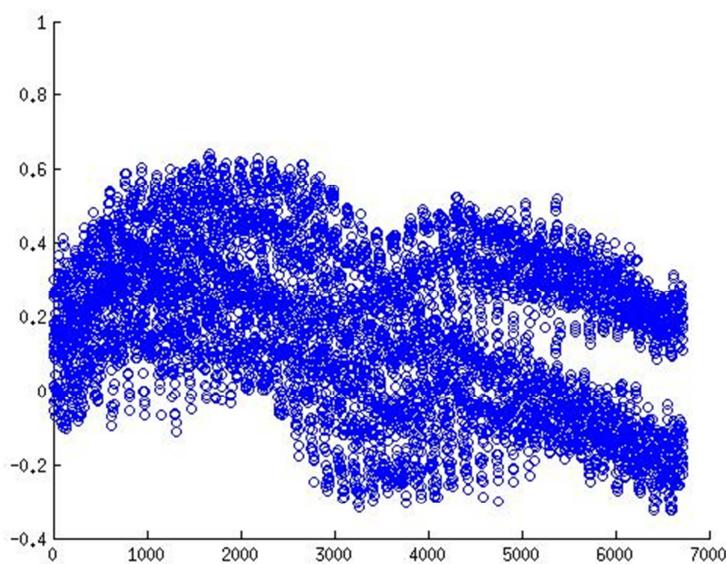


Figura 63: nube de puntos del resultado de la función *fitness* IS – Experimento 10, 11 y 12

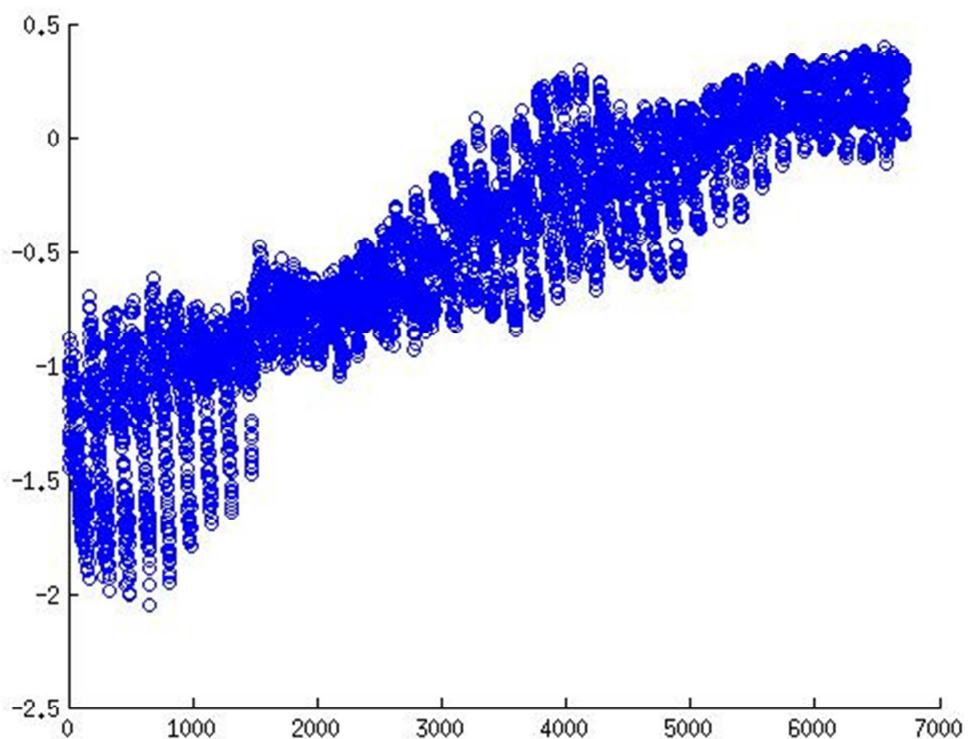


Figura 64: nube de puntos del resultado de la función *fitness* OOS – Experimento 10, 11 y 12

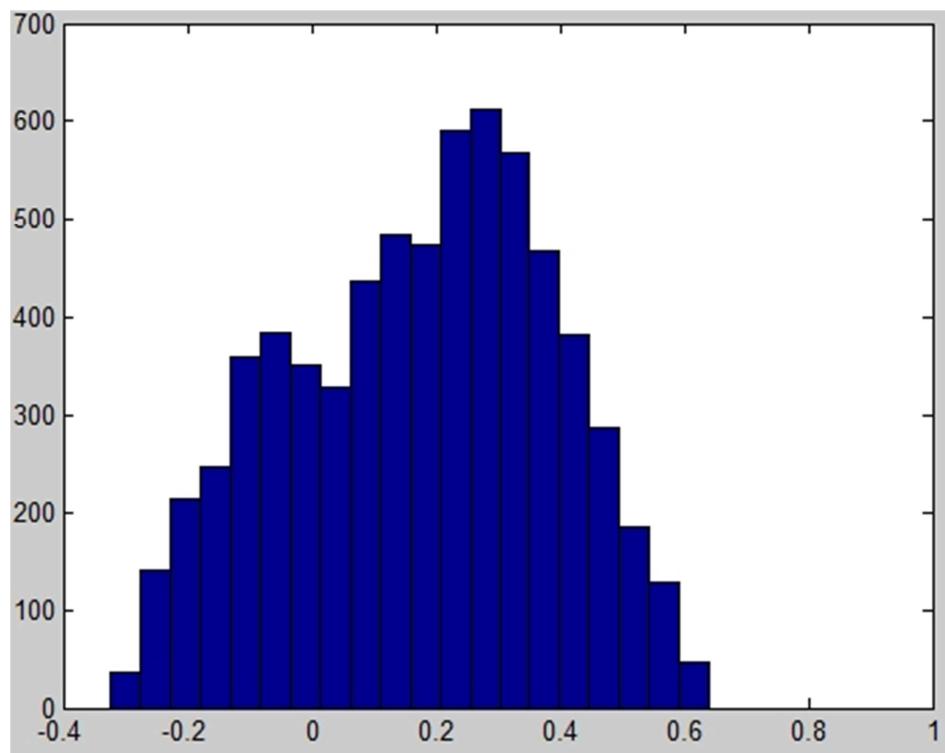


Figura 65: histograma de los valores de la función *fitness* IS – Experimento 10, 11 y 12

periSMA	periEMA	hComienzo	hFin	Coeficiente de Determinación
200	100	0900	2200	0.5222

Tabla 7: mejor combinación de parámetros – Experimento 10, 11 y 12

Retorno IS	Máximo drawdown IS	Función fitness IS	Retorno OOS	Máximo drawdown OOS	Función fitness OOS
9621	1987.5	0.3010	-8275.5	12594	-0.998

Tabla 8: retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 10, 11 y 12

Número total de trades	Trades en tramo alcista	Trades en tramo bajista	Trades en tramo lateral
995	587	242	166

Tabla 9: *trades* en tramo alcista | bajista | lateral – Experimento 10, 11 y 12

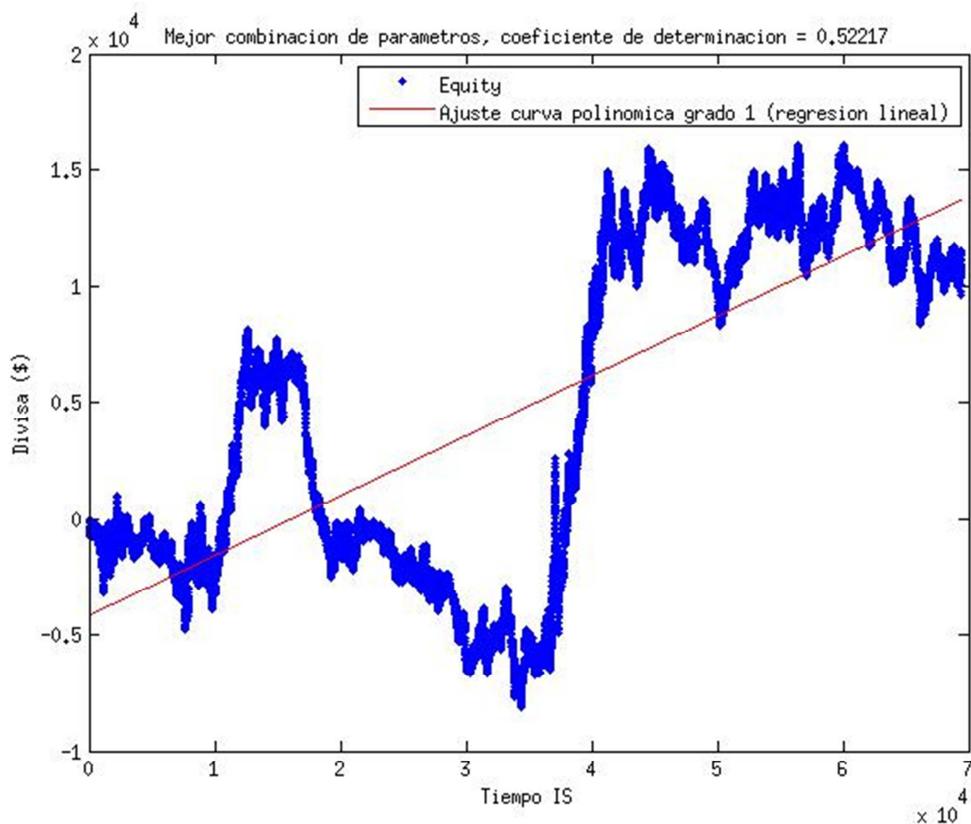


Figura 66: *equity* IS – Experimento 10, 11 y 12

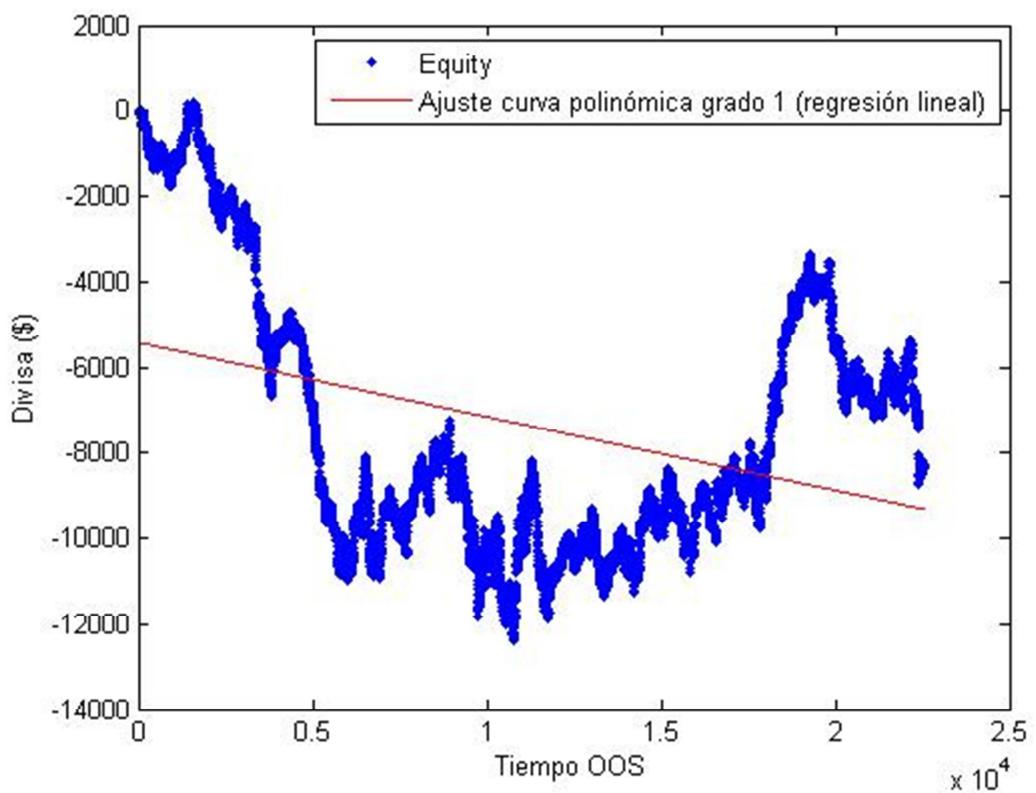


Figura 67: *equity* OOS – Experimento 10, 11 y 12

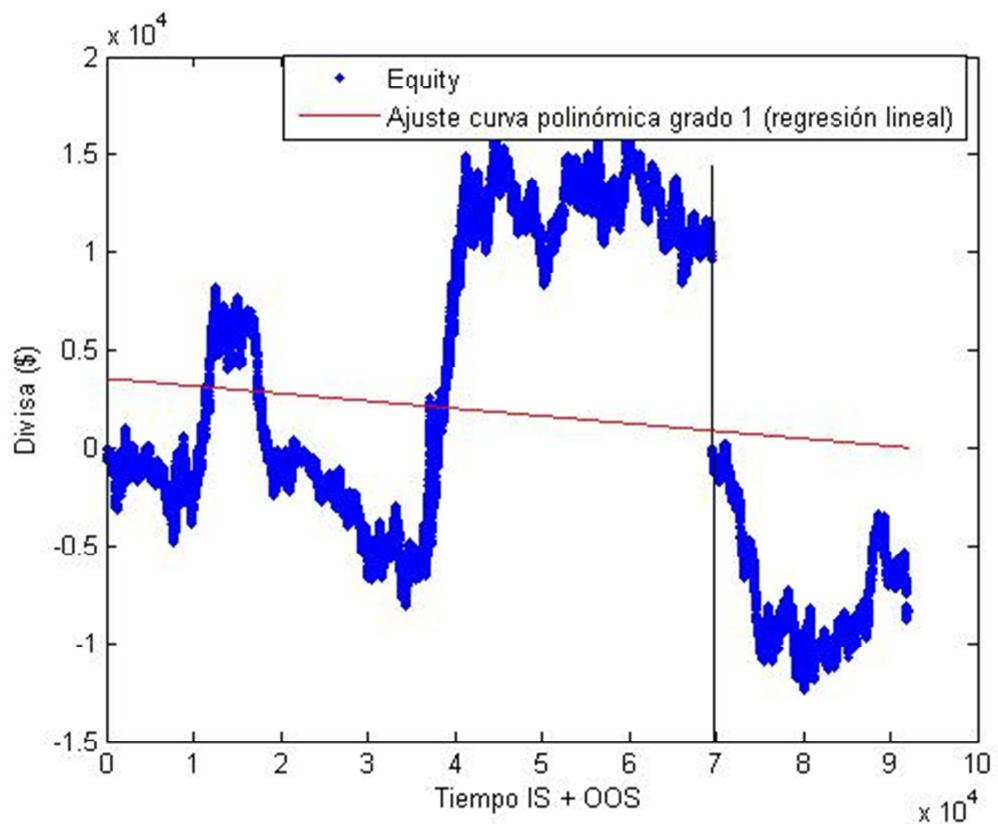


Figura 68: *equity* IS + OOS – Experimento 10, 11 y 12

Nombre del experimento: Experimento 13, 14 y 15

Número de combinaciones: 10404

Función *fitness*: SQN

Filtro: 50, 150 y 300

Tiempo de ejecución de la optimización (paralelo) = 17.26 minutos

Tiempo de ejecución de la optimización (no paralelo) = 18.16 minutos

Mem[|||||100.0%] 8058/24089MB]

Figura 69: consumo RAM – Experimento 13, 14 y 15

Como mínimo necesitamos una variable que ocupa $69441 * 10404 * 8 = 5.78$ GB

1 [|||||100.0%] 4 [|||||100.0%] 7 [| | 0.9%] 10 [| 0.5%]
2 [|||||100.0%] 5 [| | 1.9%] 8 [| 0.0%] 11 [| | 2.4%]
3 [|||||100.0%] 6 [| 0.0%] 9 [|||||100.0%] 12 [|||||100.0%]

Figura 70: carga del sistema – 6 cores – Experimento 13, 14 y 15

Load average: 6.01 5.66 3.58

Figura 71: carga media del sistema – Experimento 13, 14 y 15

PBO = 129/252 (51.2 %)

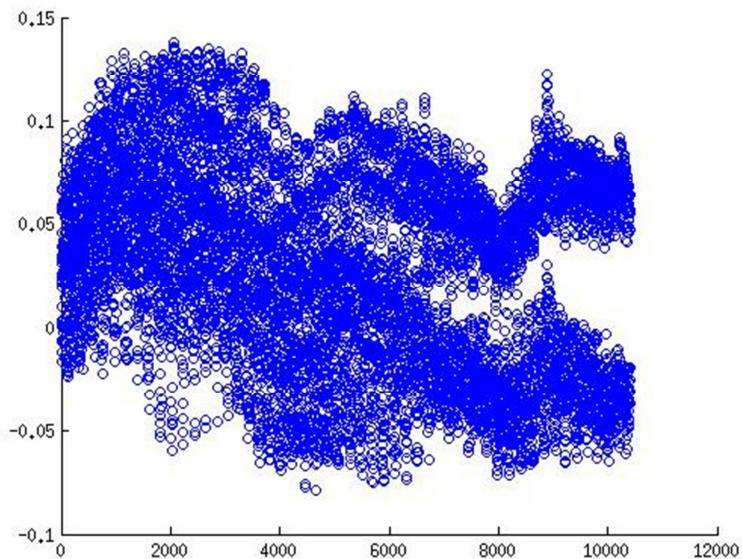


Figura 72: nube de puntos del resultado de la función *fitness* IS – Experimento 13, 14 y 15

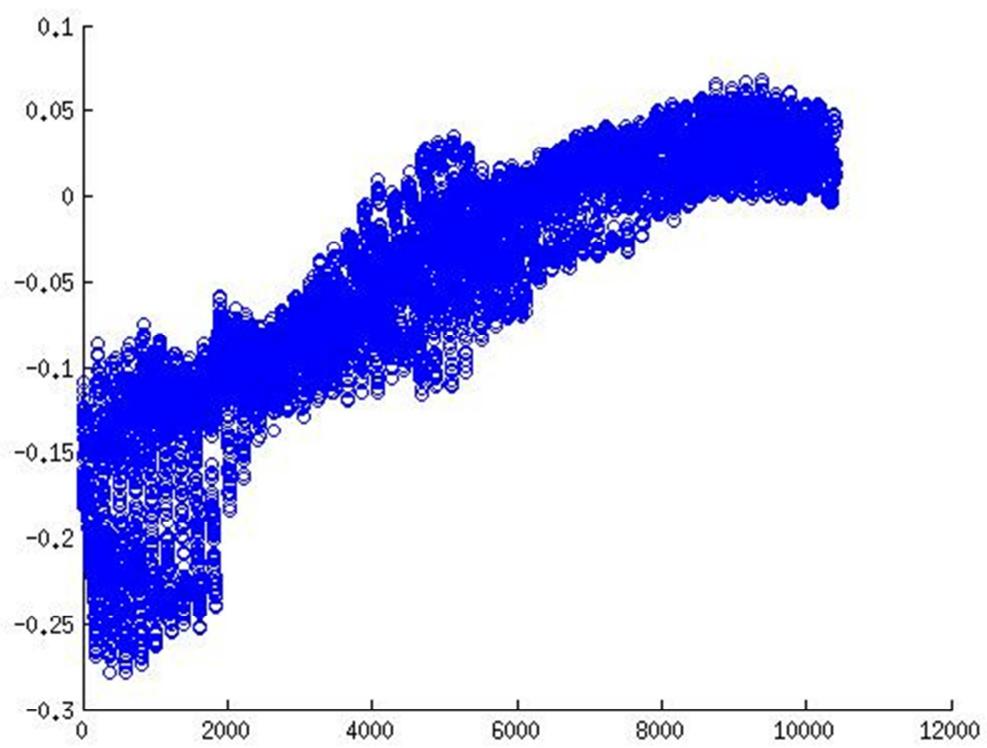


Figura 73: nube de puntos del resultado de la función *fitness* OOS – Experimento 13, 14 y 15

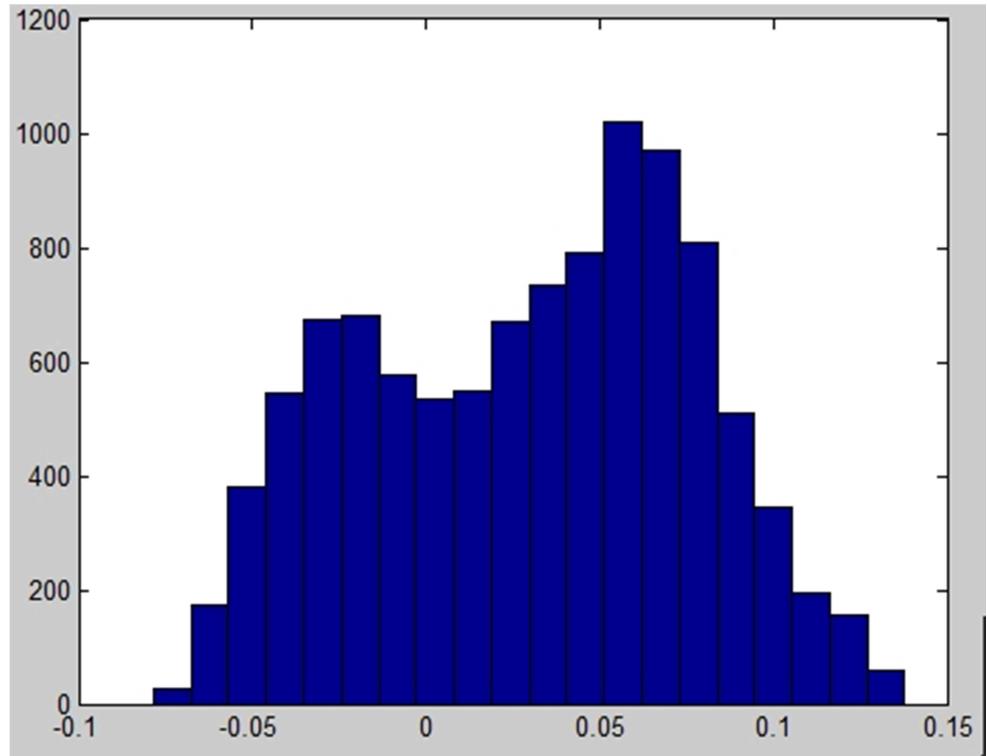


Figura 74: histograma de los valores de la función *fitness* IS – Experimento 13, 14 y 15

periSMA	periEMA	hComienzo	hFin	Coeficiente de Determinación
200	120	0900	2200	0.5108

Tabla 10: mejor combinación de parámetros – Experimento 13, 14 y 15

Retorno IS	Máximo drawdown IS	Función fitness IS	Retorno OOS	Máximo drawdown OOS	Función fitness OOS
8757.4	1987.5	0.0608	-11618	14754	-0.171

Tabla 11: retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 13, 14 y 15

Número total de trades	Trades en tramo alcista	Trades en tramo bajista	Trades en tramo lateral
1003	590	246	167

Tabla 12: *trades* en tramo alcista | bajista | lateral – Experimento 13, 14 y 15

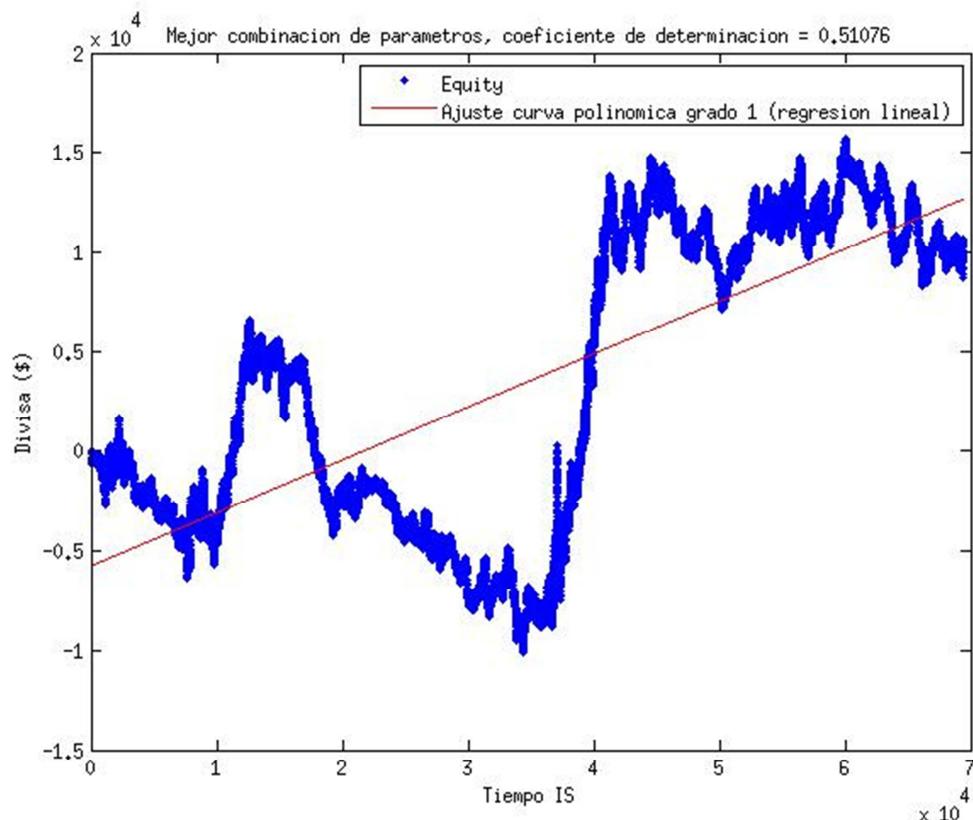


Figura 75: *equity* IS – Experimento 13, 14 y 15

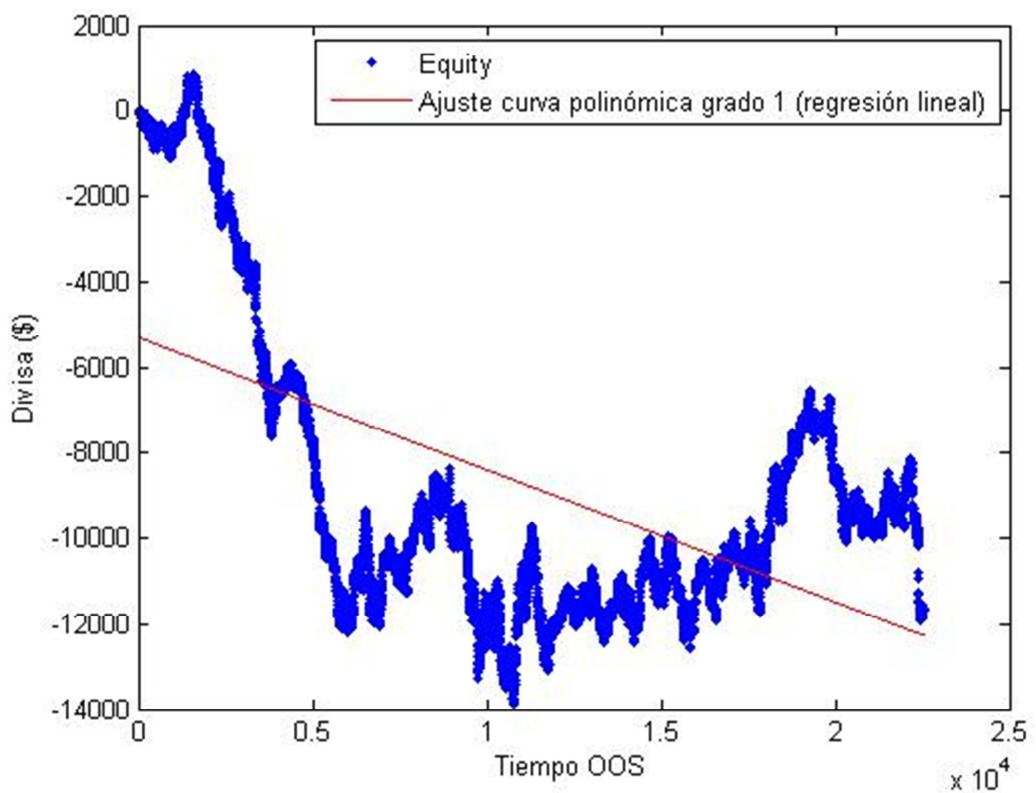


Figura 76: *equity* OOS – Experimento 13, 14 y 15

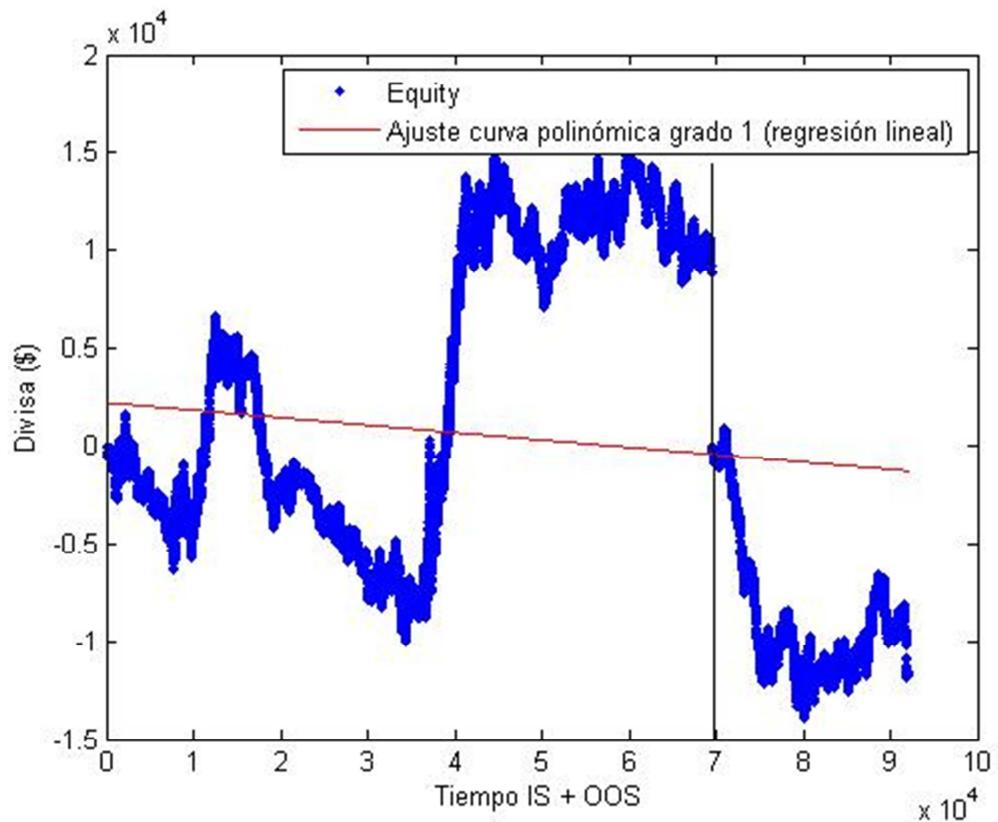


Figura 77: *equity* IS + OOS – Experimento 13, 14 y 15

4.2.1 Probabilidad de sobreoptimización

En el presente experimento voy a emplear el rango fijo de 10404 combinaciones. Los resultados serán expuestos de la siguiente manera:

- Número de combinaciones. Muestreando las 10404 combinaciones.
- Probabilidad de sobreoptimización (PBO)
- Retorno IS
- Retorno OOS

El rango seleccionado es:

periSMA 100:1:150
periEMA 50:1:100
hComienzo 0900:100:1000
hfin 2100:100:2200

Comenzamos con 10404 combinaciones:

Probabilidad de sobreoptimización (PBO) = 149/252 (59.13 %)

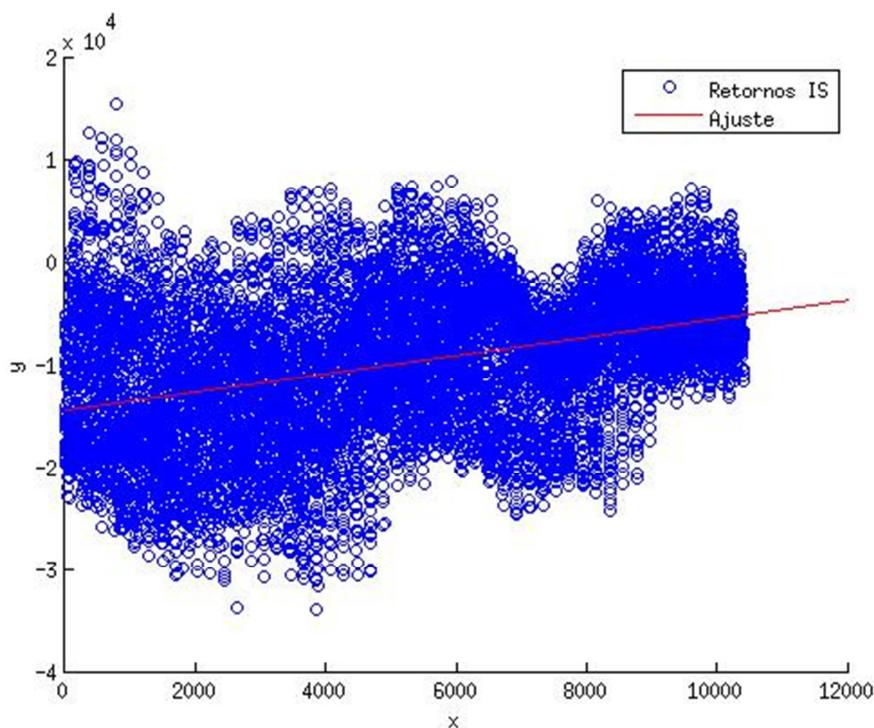


Figura 78: retornos IS – rango fijo de parámetros – 10404 combinaciones

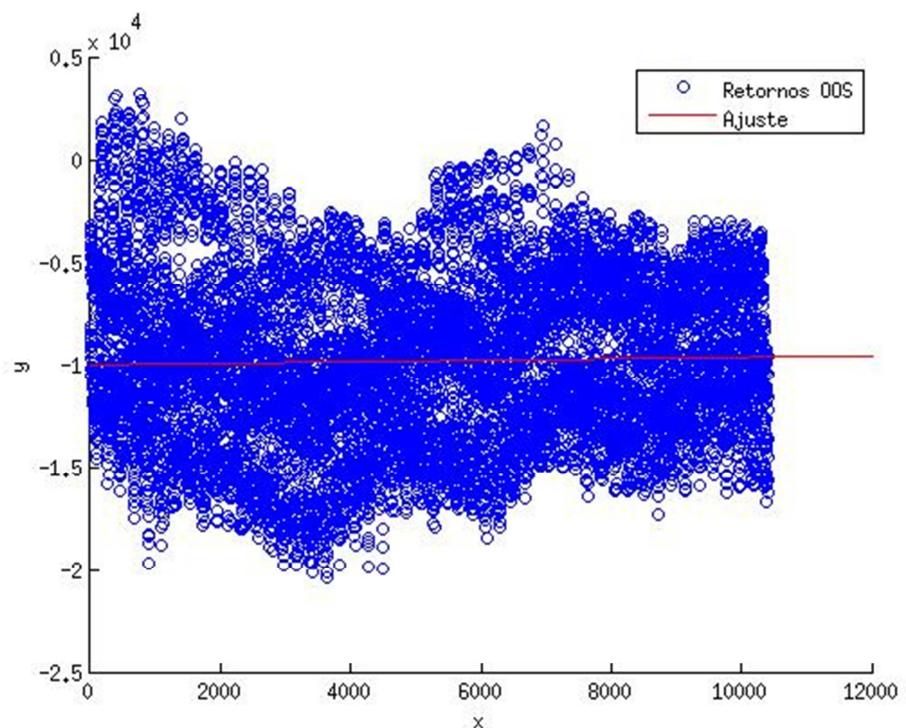


Figura 79: retornos OOS – rango fijo de parámetros – 10404 combinaciones

Muestreo de dos en dos respecto la anterior muestra, 5202 combinaciones:

Probabilidad de sobreoptimización (PBO) = 127/252 (50.4 %)

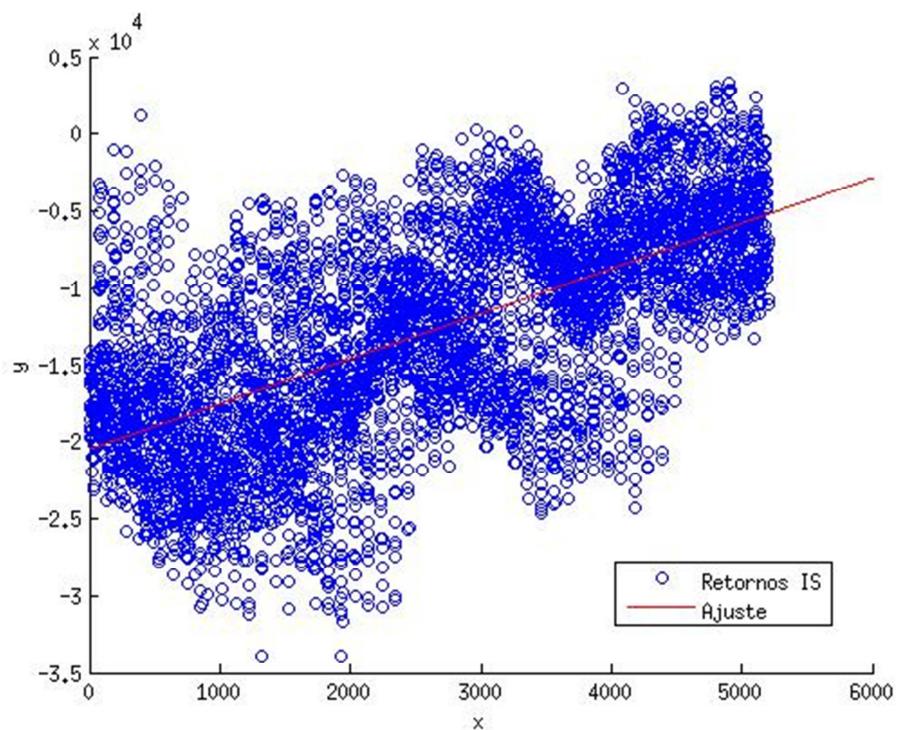


Figura 80: retornos IS – rango fijo de parámetros – 5202 combinaciones

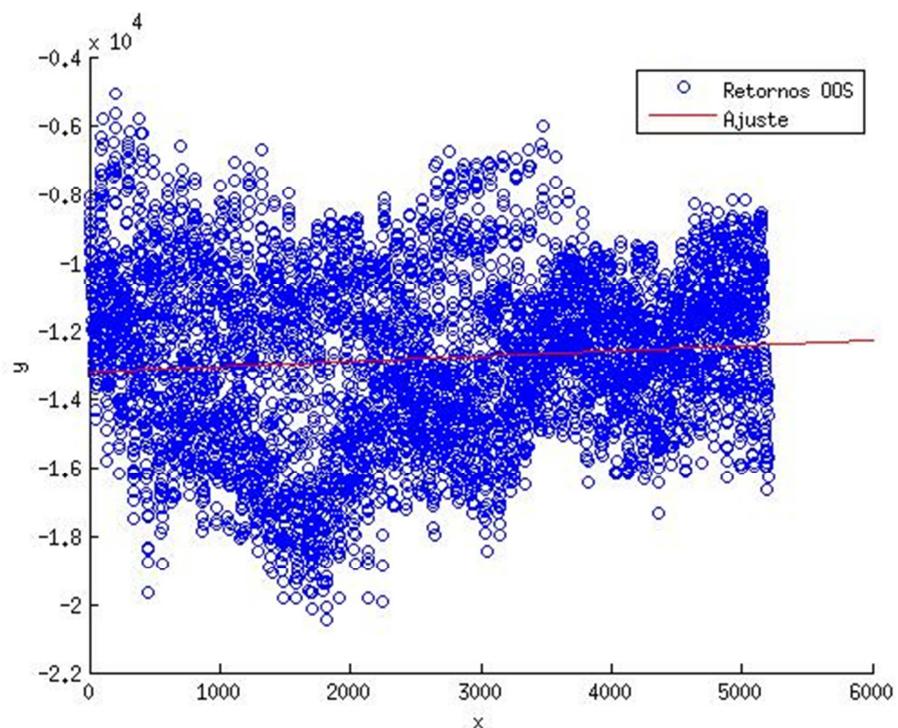


Figura 81: retornos OOS – rango fijo de parámetros – 5202 combinaciones

Muestreo de dos en dos respecto la anterior muestra, 2601 combinaciones:

Probabilidad de sobreoptimización (PBO) = 120/252 (47.62 %)

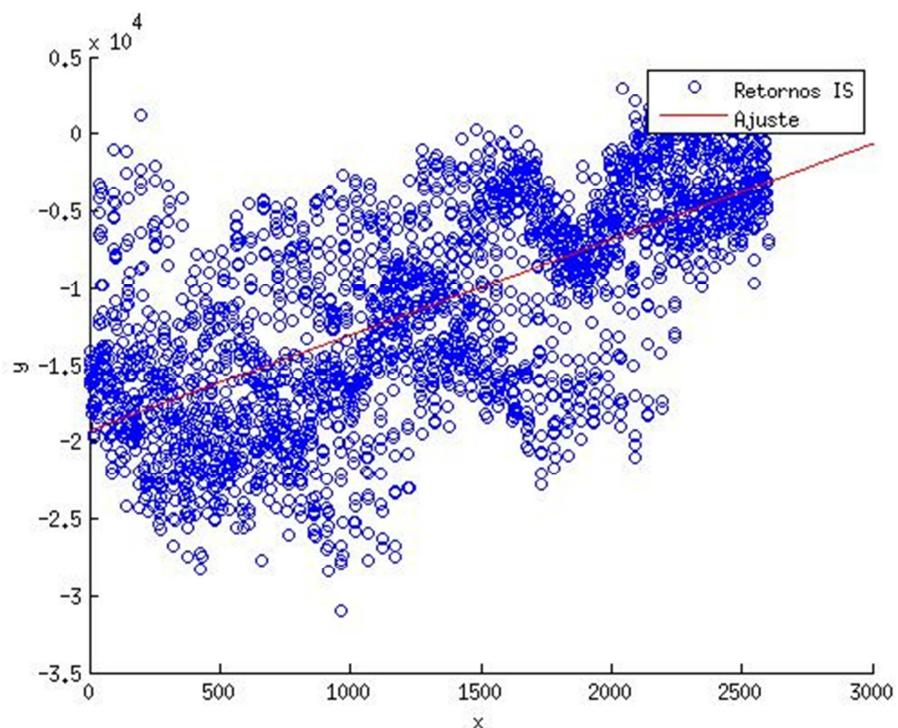


Figura 82: retornos IS – rango fijo de parámetros – 2601 combinaciones

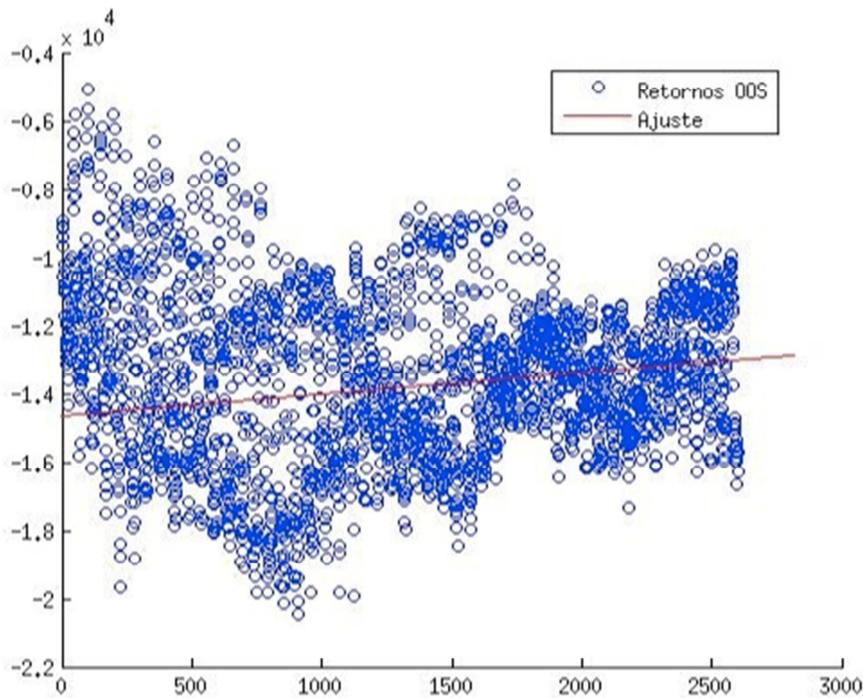


Figura 83: retornos OOS – rango fijo de parámetros – 2601 combinaciones

4.3. Fase 3

Nombre del experimento: Experimento 1, 2 y 3

Número de combinaciones: 1764

Función *fitness*: SQN

Filtro: 50, 150 y 300

Retorno OOS	Máximo drawdown IS	Máximo drawdown OOS	Máximo drawdown con Monte Carlo
-11262	1975	12610	17567

Tabla 13: retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 1, 2 y 3

Ratio Calmar;

$$CR = \frac{\text{Retorno OOS}}{\text{Máximo drawdown con Monte Carlo}} = \frac{-11262}{17567} = -0.64$$

Nombre del experimento: Experimento 10, 11 y 12
 Número de combinaciones: 6724
 Función *fitness*: SR
 Filtro: 50, 150 y 300

Retorno OOS	Máximo drawdown IS	Máximo drawdown OOS	Máximo drawdown con Monte Carlo
-8275.5	1987.5	12594	18836

Tabla 14: retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 10, 11 y 12

Ratio Calmar;

$$CR = \frac{\text{Retorno OOS}}{\text{Máximo drawdown con Monte Carlo}} = \frac{-8275.5}{18836} = -0.43$$

Nombre del experimento: Experimento 13, 14 y 15
 Número de combinaciones: 10404
 Función *fitness*: SQN
 Filtro: 50, 150 y 300

Retorno OOS	Máximo drawdown IS	Máximo drawdown OOS	Máximo drawdown con Monte Carlo
-11618	1987.5	14754	20743

Tabla 15: retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 13, 14 y 15

Ratio Calmar;

$$CR = \frac{\text{Retorno OOS}}{\text{Máximo drawdown con Monte Carlo}} = \frac{-11618}{20743} = -0.56$$

4.4. Experimentos número veinte y veintiuno – sin contar con fase 1

Tomamos en cuenta que el histórico continuo filtrado y muestreado en 15 minutos ocupa en un vector fila 69432 columnas. Se estará preguntando, ¿Por qué no ocupa lo mismo que el histórico continuo con la fase 1 aplicada? Debido al cambio de *rollover* aplicado. Como en la fase 1 pudimos identificar que la negociación cambia una semana antes al tercer viernes del mes oficial, ésto significa que la negociación disminuye la última semana de vencimiento oficial y por eso tendremos menos datos (barras). Realizo estos experimentos extras a modo de demostración para ser consciente de cómo puede llegar a distorsionarse el resultado estimado durante el análisis.

Recordamos que la mejor combinación de parámetros es:

periSMA	periEMA	hComienzo	hFin	Coeficiente de Determinación
200	100	0900	2200	0.5222

Tabla 16: mejor combinación de parámetros

Respecto el histórico de proveedor1 sin realizar la fase 1 del Proceso ATurel

Muestra	Columnas	Sharpe	SQN	Beneficio	Max. drawdown	Trades
IS	68863	0.0405	0.009	1290.9	18783	998

Tabla 17: datos IS – sin fase 1 – proveedor1

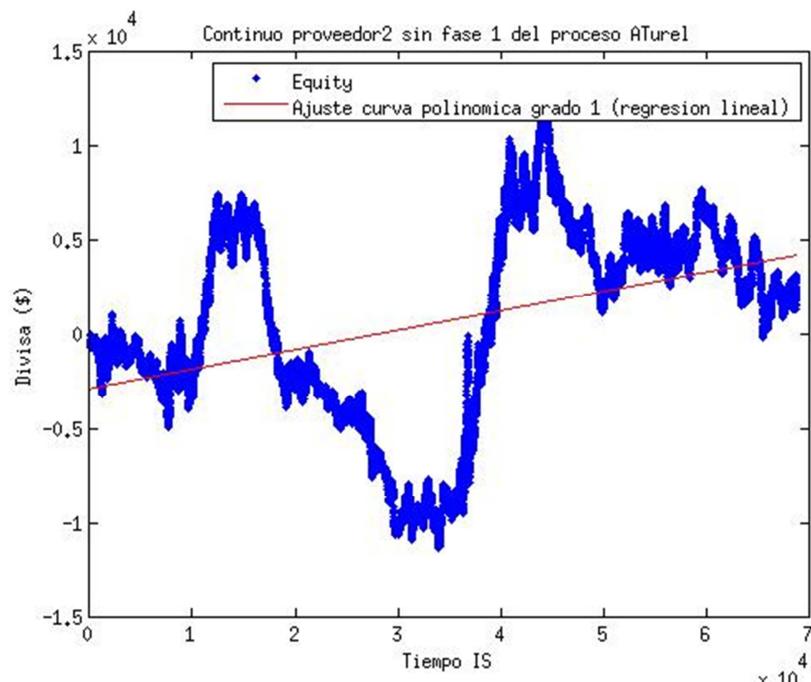


Figura 84: equity IS – sin fase 1 – proveedor1

Muestra	Columnas	Sharpe	SQN	Beneficio	Max. drawdown	Trades
OOS	22381	-1.0025	-0.1206	-8256.2	12970	324

Tabla 18: datos OOS – sin fase 1 – proveedor1

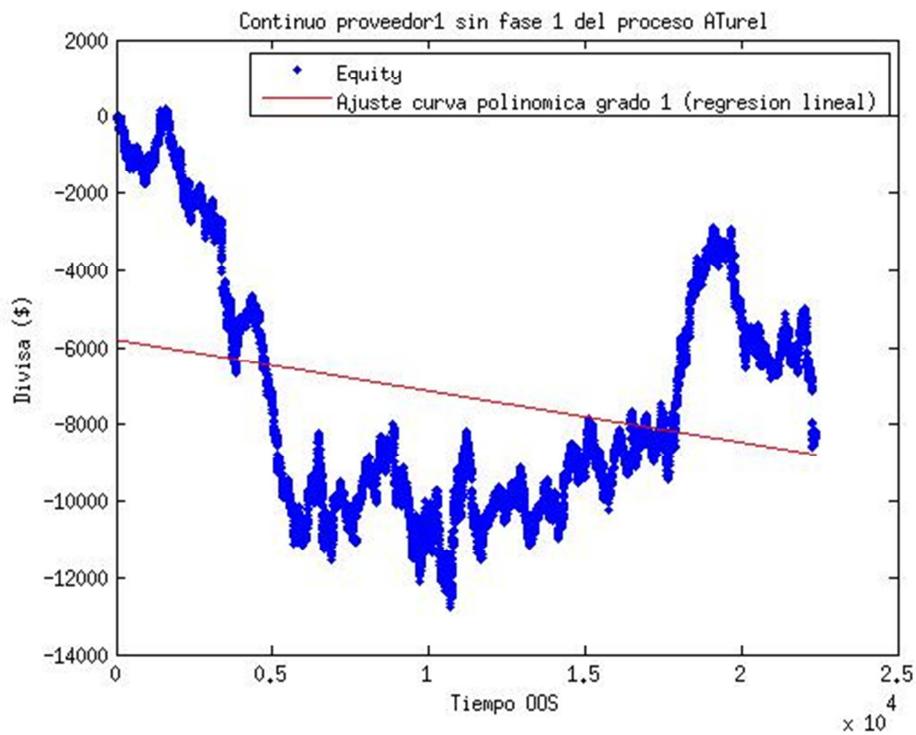


Figura 85: equity OOS – sin fase 1 – proveedor1

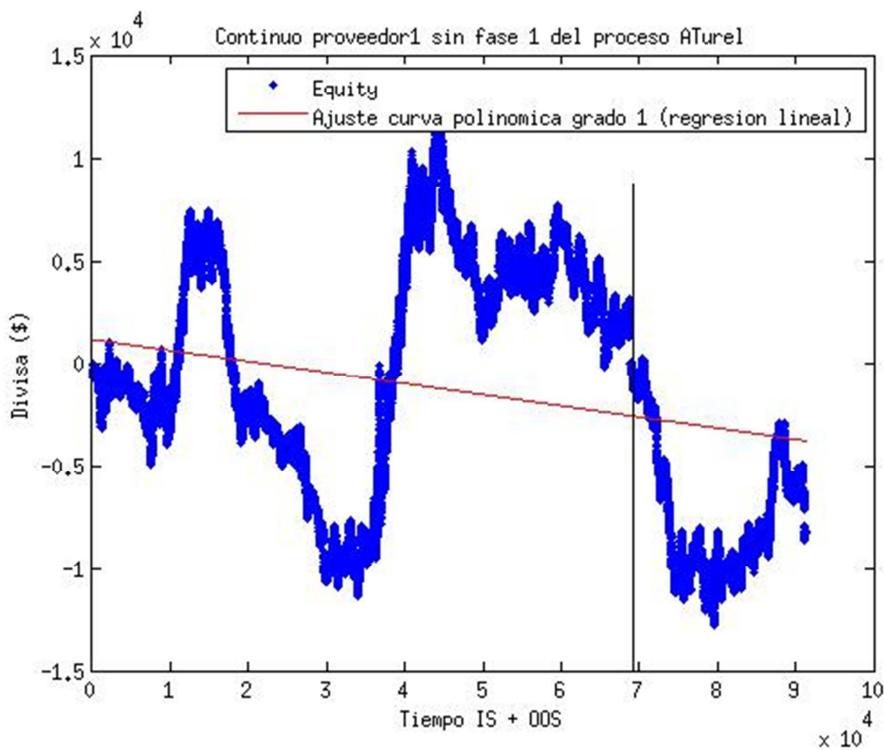


Figura 86: equity IS + OOS – sin fase 1 – proveedor1

Respecto el histórico de proveedor2 sin realizar la fase 1 del Proceso ATurel

Muestra	Columnas	Sharpe	SQN	Beneficio	Max. drawdown	Trades
IS	69390	-0.067	-0.0149	-2168.2	17317	996

Tabla 19: datos IS – sin fase 1 – proveedor2

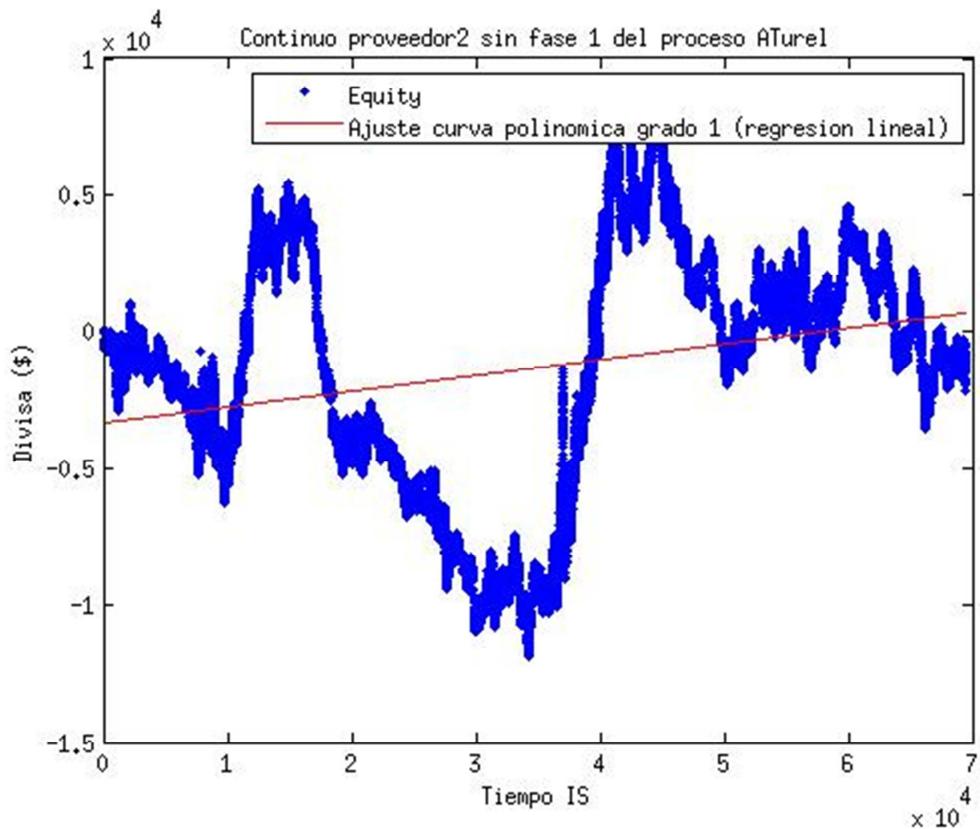


Figura 87: equity IS – sin fase 1 – proveedor2

Muestra	Columnas	Sharpe	SQN	Beneficio	Max. drawdown	Trades
OOS	22441	-0.8885	-0.1097	-7463.3	13297	325

Tabla 20: datos OOS – sin fase 1 – proveedor2

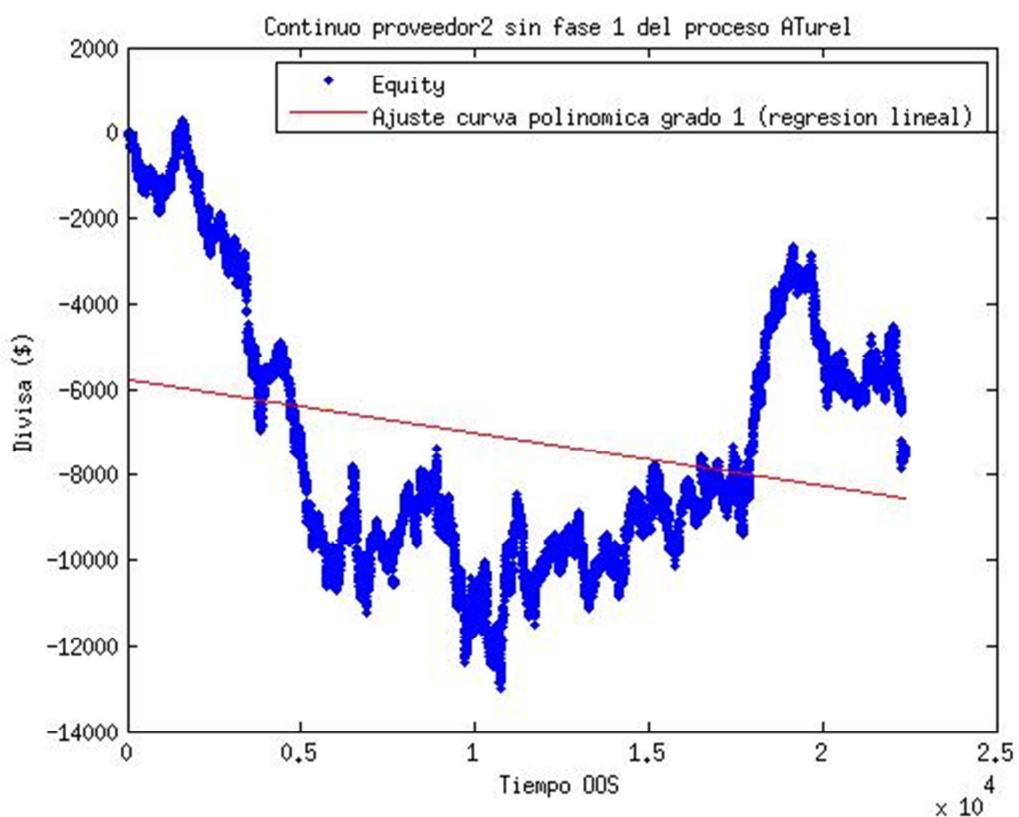


Figura 88: *equity OOS – sin fase 1 – proveedor2*

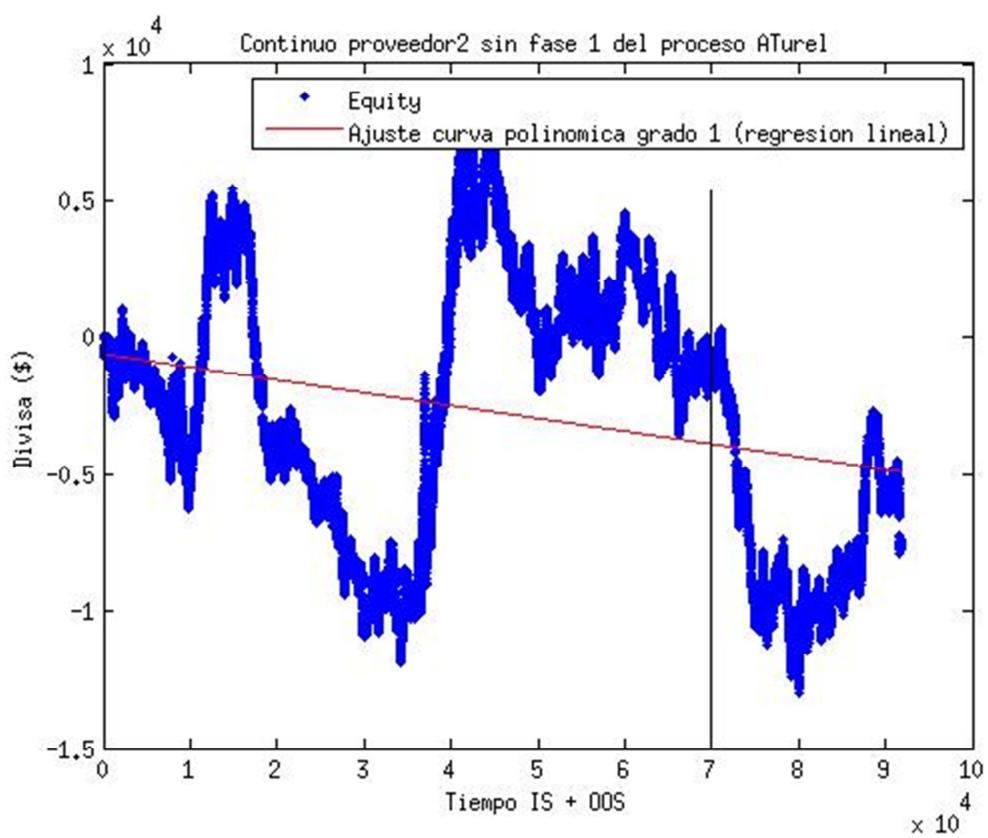


Figura 89: *equity OOS – sin fase 1 – proveedor2*

4.5. Experimento número veintidós – medidas computacionales

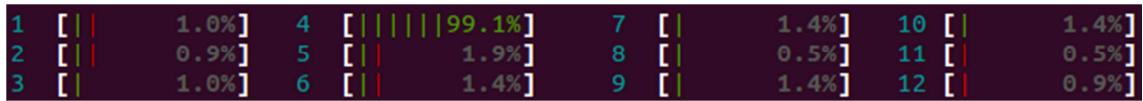


Figura 90: carga del sistema – no paralelo – Experimento 13, 14 y 15

Load average: 1.24 2.66 3.60

Figura 91: carga media del sistema – no paralelo – Experimento 13, 14 y 15

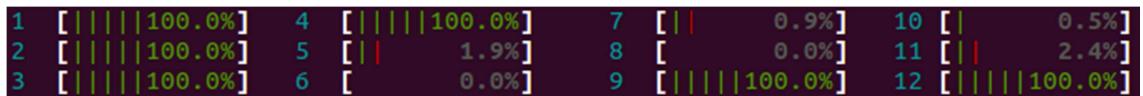


Figura 92: carga del sistema – 6 cores – Experimento 13, 14 y 15

Load average: 6.01 5.66 3.58

Figura 93: carga media del sistema – Experimento 13, 14 y 15

Es normal que la carga media aumente cuando utilizamos más cores del sistema en el caso paralelo. Si la carga media fuese entorno a 12 querría decir que se está utilizando casi toda la potencia disponible por el sistema (próximo al 100%).

	1762 combinaciones	6724 combinaciones	10404 combinaciones
Tiempo no paralelo (minutos)	3.09	12.31	18.16
Tiempo en paralelo (minutos)	2.76	11.86	17.26
Tiempo ganado (%)	10.68	3.66	7.2

Tabla 21: tiempos de ejecución

Como se puede observar no he ganado un rendimiento computacional apreciable.

Veamos ahora qué ocurriría si no reservo memoria en las variables:

Tiempo de ejecución de la optimización (no paralelo y reservando memoria),

18.16 minutos

Tiempo de ejecución de la optimización (no paralelo y sin reservar memoria),

73.42 minutos

Gracias a la reserva de memoria, la ejecución ha sido > 4 veces más rápida.

5. Conclusiones y trabajos futuros

Son muchos los resultados que se han obtenido y todos ellos, como en cualquier otro estudio realizado, nos tienen que resolver los objetivos marcados en el punto 1.6 a modo de veredictos. Lógicamente, este trabajo pretende hacerse un hueco entre toda la investigación que hay en estos momentos sobre Algo Trading y análisis cuantitativo y por ello qué mejor noticia que resulte provechoso para continuar y avanzar con nuevas líneas de trabajo para aquellas personas que deseen comenzar en este fascinante mundo de los Sistemas Algorítmicos y el análisis cuantitativo. Me gusta más la práctica que la teoría pero claro, como ya cité, no es recomendable invertir por simple curiosidad sin conocer previamente su metodología. Con este trabajo se pretende crear, si Dios quiere, nuevas líneas de investigación y animar para seguir mejorando el procedimiento de estudio, el Proceso ATurel. Pongamos por caso que a lo mejor se podrá añadir a la fase tres más elementos de control de riesgo, por ejemplo el famoso VaR y CVaR, ampliamente utilizado en el sector financiero y considerado ya como si de un estándar se tratase.

5.1. Resumen del trabajo

Comenzamos con el objetivo 1;

En el punto 4.2.1 de la página 88, el lector podrá comprobar que el experimento elaborado demuestra lo que planteaba en un principio, "...independientemente de la función *fitness* empleada, se produce sobreoptimización si optimizamos el Sistema Algorítmico de *Trading* con un elevado número de combinaciones...". El rango fijo de combinaciones seleccionado, unas 10404 combinaciones, da como resultado una probabilidad de overfitting (59.13%) mayor que si muestreo para obtener submatrices con menor número de combinaciones. Si muestreo una vez de dos en dos (5202 combinaciones), la probabilidad de sobreoptimización es del 50.4% y si vuelvo a muestrear de dos en dos (2601 combinaciones), la probabilidad de sobreoptimización es del 47.62%. Las probabilidades de sobreoptimización son aparentemente altas porque el parámetro *s* está fijado a 10 y ello da lugar a 252 posibles candidatos entre 10404 combinaciones. Se trata de un parámetro que conviene optimizar en un trabajo especializado en la probabilidad de sobreoptimización. Por otro lado las probabilidades no están significativamente separadas y ello me hace sospechar en los retornos representados, tanto IS como OOS. Resulta que el rendimiento OOS aumenta de forma débil entre submatrices pero se aprecia que mejora a medida que muestreamos para obtener submatrices con menor número de combinaciones. Lo puede ver con el ajuste realizado a la nube de puntos por regresión lineal.

Pasando a comentar los resultados referidos al objetivo 2;

El rango de parámetros seleccionado es distinto al anterior para mostrar rendimientos no especialmente buenos. Se trata de un trabajo que trata de expresar una metodología de simulador para analizar correctamente y para demostrar los objetivos se ha utilizado un Sistema Algorítmico de Trading de tantos que existen actualmente.

Pero el objetivo no es demostrar que se trata de un buen Sistema Algorítmico de Trading.

Después de realizar los dieciocho experimentos de la página 65, con un Ratio Calmar de -0.43 he obtenido como mejor combinación de parámetros;

$$\text{periSMA} = 200 \quad \text{periEMA} = 100 \quad \text{hComienzo} = 0900 \quad \text{hFin} = 2200$$

En este caso el Sharpe Ratio ha sido el que más rendimiento OOS nos proporciona, -8275.5 dólares, en función del riesgo que estamos asumiendo con el máximo *drawdown* (CR). También es verdad que el resto de experimentos (con otro número de combinaciones) no hace mejorar dicho ratio respecto el SQN y que resultan ser iguales en los dos últimos experimentos (14-18). Por tanto no aprecio un claro indicio de que un ratio te proporcione más rendimiento que otro porque si la preocupación está en trabajar con una muestra significativa, por ejemplo un alto número de *trades*, se pueden incorporar los filtros mencionados en la página 60 para cualquier función *fitness* a optimizar. Un comentario importante sobre la representación de la función *fitness* OOS es sobre el claro agotamiento de la misma a medida de aumentamos el número de combinaciones. La representación gráfica de la página 146 es creciente mientras que el de la página 168 muestra indicios de agotamiento y descenso. Esto nos puede dar otra pista sobre la problemática de la sobreoptimización, viendo que empeora a medida que aumentamos el número de combinaciones.

Por otro lado, en el objetivo 2 bis queda demostrado desde la página 170 que el máximo *drawdown* aumenta después de aplicar un método Monte Carlo. Además era de suponer que el máximo *drawdown* IS sería menor que el de la muestra OOS pero siempre puede haber excepciones y por eso aplico toda la muestra (IS+OOS) a la simulación Monte Carlo. Puntualizar que otro de los temas importantes en este campo es conocer la cartera inicial necesaria para comenzar la inversión. Una de las principales variables es el máximo *draw down* y generalmente cada inversor aplica su propia regla añadiendo 'factores de salvaguarda'. Un ejemplo podría ser;
Cartera inicial = (garantías exigibles + máximo drawdown) * factor.

Pasamos al objetivo 3 donde pretendo demostrar tanto la importancia de realizar un preprocessado de datos para filtrar posibles irregularidades como también la importancia de unir correctamente vencimientos al tratarse en este caso de una simulación con el derivado futuro. En general los históricos de datos eran correctos con la excepción de que el proveedor2 ha presentado datos algunos fines de semana que no debiera haber mercado y también datos al finalizar el vencimiento oficial.

Después de realizar la metodología de construcción del histórico continuo ha quedado demostrado, gracias al estudio del volumen de negociación, que el cambio de vencimiento se debe realizar una semana antes de la semana oficial (3º viernes de cada mes) porque el volumen aumenta en dicha semana que atañe al cambio de vencimiento.

El hecho de haber filtrado y construido los históricos financieros desde cero nos debe dar mayor seguridad de los resultados que se vayan a obtener en las sucesivas

simulaciones. Sin ir más lejos, comentaba al principio que la mejor combinación de parámetros refleja un rendimiento OOS de -8275.5 dólares mientras que si aplicamos el *backtest* al histórico sin tratar del proveedor1 obtenemos -8256.2 dólares y si realizamos el mismo proceso con el proveedor2 obtenemos -7463.3 dólares. A simple vista no aparenta grandes diferencias pero si el error se arrastra a sucesivos estudios podría ser catastrófico por estar trabajando con una muestra no tratada.

El último punto a comentar del objetivo 3 es el procedimiento que me he inventado para distinguir los tramos alcistas, bajistas y laterales. En los resultados de la fase 2 se aprecia que la mayoría de los *trades* se realizan en tendencia, de ahí que al Sistema Algorítmico de Trading empleado para los experimentos se le considere un sistema tendencial. En concreto, el potencial de los *trades* se realizan en tendencia alcista. Aun así se trata de un comienzo de investigación donde se puede perfeccionar y como principal objetivo saber si el sistema se comporta mejor en tramos alcistas | bajistas o laterales para que la elección de *portfolio* de sistemas esté lo más incorrelado posible.

Por último y no menos importante expresamos las conclusiones computacionales. Para procesar todos los datos (*equity*, función *fitness*, número de *trades* y rendimientos en general) de simulación se ha podido comprobar que se necesita sobre todo memoria RAM para almacenar matrices de datos tan pesadas y eso que la más grande con la que he trabajado ha sido con un histórico de 3 años, *timeframe* de 15 minutos y 10404 combinaciones (5.78 GB como mínimo). En cuanto a los retardos de ejecución, la utilización de 6 *cores* en vez de uno no se aprecia grandes diferencias, esto me hace sospechar en la plataforma MATLAB porque si se hubiera hecho por ejemplo en C o JAVA, estoy seguro que las diferencias si serían significativas. En cambio, el hecho de reservar memoria en las variables necesarias, previo a la ejecución, nos ha aportado ganar más de cuatro veces el tiempo de ejecución en el caso de no haberlo hecho.

5.2. Contribuciones del trabajo

El presente trabajo va a poder contribuir con el proyecto ROBOTRADER para estructurar cada vez mejor el entrenamiento y gestión de puntuación de su competición y como decíamos al comienzo del punto 5, como semilla, entre los pocos trabajos relacionados con el Algo Trading, para que aumente la investigación en universidades de carácter técnico principalmente. Es una materia que aun no siendo muy conocida, cuando se explica en alguna conferencia o evento parecido, el interés surge por casi todos los alumnos asistentes.

5.3. Trabajos futuros

El lector ha podido notar la complejidad de este campo. Demostrar que los procesos implementados son viables económicamente hablando, implica realizar muchas más simulaciones y estudios para que el dinero invertido esté en consonancia con la rentabilidad esperada y el riesgo asumido. El Proceso ATurel le queda más rodaje por analizar tanto inter-fase como intra-fase. Por ejemplo, se debería añadir el estudio de *portfolios*, cartera inicial y gestión monetaria.

6. Referencias

- Trading Systems and Methods – Perry J. Kaufman. Fifth Edition, Wiley
- Sistemas de Gestión del Capital en Sistemas Automáticos de Inversión. Diseño e implementación. – PFC ETSIT-UPM – Ignacio Cruz Caridad
- THE PROBABILITY OF BACKTEST OVERFITTING 2013, David H. Bailey, Jonathan M. Borwein, Marcos López de Prado, Qiji Jim Zhu
- PSEUDO-MATHEMATICS AND FINANCIAL CHARLATANISM: THE EFFECTS OF BACKTEST OVERFITTING ON OUT-OF-SAMPLE PERFORMANCE 2013, David H. Bailey, Jonathan M. Borwein, Marcos López de Prado, Qiji Jim Zhu
- Toolbox de Matlab, <http://www.mathworks.es/> última consulta realizada el 01/12/2014
- C89 / C99 standard, <http://port70.net/~nsz/c/> última consulta realizada el 01/12/2014
- Alexander Elder Quotes,
https://www.goodreads.com/author/quotes/52854.Alexander_Elder (punto 1.), última consulta realizada el 01/12/2014
- 2010 Flash Crash http://en.wikipedia.org/wiki/2010_Flash_Crash (punto 1.2), última consulta realizada el 01/12/2014
- Personal computer http://en.wikipedia.org/wiki/Personal_computer#History (punto 1.2), última consulta realizada el 01/12/2014
- Historia del Trading <http://www.etoro.com/blog/market-news-es/17102013/historia-del-trading/?lang=es> (punto 1.2), última consulta realizada el 01/12/2014
- Mercados Financieros & Computer Science <http://www.robotraderworld.com>) (punto 1.3), última consulta realizada el 01/12/2014
- Apuntes de la asignatura; Laboratorio de Técnicas de Soporte a la Decisión (LTSD) en E.T.S.I. Telecomunicación (punto 1.4.1 y 1.4.2)
- Medidas de dispersión http://es.wikipedia.org/wiki/Medidas_de_dispersión (punto 1.4.2), última consulta realizada el 01/12/2014
- Instituto Nacional de Estadística <http://www.ine.es> (punto 1.4.2), última consulta realizada el 01/12/2014

- "Equity Curve" a fondo

http://www.tradingsys.org/index.php?option=com_content&task=view&id=90&Itemid=50 (punto 1.4.2), última consulta realizada el 01/12/2014

- Improving System Models Using Monte Carlo Techniques on Plant Models

http://www.mathworks.com/tagteam/52422_AIAA-2008-7094-316.pdf (punto 1.4.3),
última consulta realizada el 01/12/2014

Conociendo el E-mini S&P 500 <http://www.pullback.es/conociendo-el-futuro-e-mini-sp-500/> (punto 2.1), última consulta realizada el 01/12/2014

- About Welles Wilder <https://www.deltasociety.com/content/about-welles-wilder> (punto 2.2), última consulta realizada el 01/12/2014

ANEXO I: Características futuro E-mini S&P 500 (CME Group)

Quotes	Settlements	Volume	Time & Sales	Contract Specs	Margins	Calendar					
Futures	Options										
Opening Date	9/9/1997										
Ticker Symbol	ES ES= Clearing View product and vendor codes										
Contract Size	\$50 x S&P 500 Index										
Tick Size (minimum fluctuation)	OUTRIGHT	0.25 index points=\$12.50									
	CALENDAR SPREAD	0.05 index points=\$2.50									
Trading Hours All time listed are Central Time	CME Globex (ETH)	MON – FRI: 5:00 p.m. previous day – 4:15 p.m.; trading halt from 3:15 p.m. – 3:30 p.m.									
Contract Months	Five months in the March Quarterly Cycle (Mar, Jun, Sep, Dec)										
Last Trade Date/Time View Calendar	CME Globex	Trading can occur up to 8:30 a.m. on the 3rd Friday of the contract month									
Settlement Procedure	Daily E-mini S&P 500 Settlement Procedure (PDF) Final E-mini S&P 500 Settlement Procedure (PDF)										
Daily Price Limits	View price limits details.										
Position Limits	See: CME Rulebook Chapter 5 - POSITION LIMIT, POSITION ACCOUNTABILITY AND REPORTABLE LEVEL TABLE										
Block Trade Eligibility	No. View more on block-trade eligible contracts.										
Block Minimum	N/A										
Rulebook Chapter	358										
Exchange Rule	These contracts are listed with, and subject to, the rules and regulations of CME.										

ANEXO II: Informe detallado - resultados fase 1

Primero comenzaremos con los resultados del Filtrado1 (por cada vencimiento por separado), después los resultados del Filtrado2 (por cada vencimiento por separado), continuaremos con los resultados sobre el período horario de máxima negociación (por cada vencimiento por separado), para así construir el histórico continuo y finalmente indicar las medidas de los tramos laterales para obtener el intervalo de confianza de la pendiente de la recta tangente para considerar o no un tramo lateral. Los resultados se expresarán en su mayoría por medio de tablas para su fácil identificación y comprensión. En base a todos los resultados obtenidos se realizarán los ajustes de borrado y por mínimos cuadrados explicados en el punto 3.1. Recordamos que los vencimientos están con *timeframe* de 1 minuto.

Comenzamos con el Filtrado1:

Vencimiento ESH10

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	94417	NO	NO	NO	0
Proveedor2	94895	NO pero cierra a las 14:30 h	NO	NO	0

Tabla 22: Filtrado 1, vencimiento ESH10

Vencimiento ESH11

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	97572	NO	NO	NO	0
Proveedor2	98209	NO pero cierra a las 14:31 h	SI, 60 líneas (desde la 92062 hasta la 92121). Pertenece al 13 de marzo de 2011	NO	0

Tabla 23: Filtrado 1, vencimiento ESH11

Vencimiento ESH12

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	95492	NO	NO	0	0
Proveedor2	96126	NO pero cierra a las 14:30 h	SI, 50 líneas (la 35991 y desde la 90335 hasta la 90383) Pertenecen al 15 de enero de 2012 y las restantes al 11 de marzo de 2012	0	0

Tabla 24: Filtrado 1, vencimiento ESH12

Vencimiento ESH13

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	89842	NO	NO	NO	0
Proveedor2	90143	NO pero cierra a las 14:30 h	SI, 55 líneas (desde la 84596 hasta la 84650) Pertenecen al 10 de marzo de 2013	NO	0

Tabla 25: Filtrado 1, vencimiento ESH13

Vencimiento ESM10

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	98340	NO	NO	NO	0
Proveedor2	98811	NO	NO	NO	0

Tabla 26: Filtrado 1, vencimiento ESM10

Vencimiento ESM11

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	98731	NO	NO	NO	0
Proveedor2	99423	NO	SI, 119 líneas (desde la 6618 hasta la 6677 y desde la 13512 hasta la 13570) Pertenece al 13 de marzo de 2011 y al 20 de marzo de 2011	NO	0

Tabla 27: Filtrado 1, vencimiento ESM11

Vencimiento ESM12

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	98552	NO	NO	NO	0
Proveedor2	99559	NO	SI, 113 líneas (desde la 5445 hasta la 5499 y desde la 12249 hasta la 12306) Pertenece al 11 de marzo de 2012 y al 18 de marzo de 2012	NO	0

Tabla 28: Filtrado 1, vencimiento ESM12

Vencimiento ESM13

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	105032	NO	NO	NO	0
Proveedor2	104985	SI, a las 15:31 h	SI, 175 líneas (desde la 5318 hasta la 5375, desde la 12104 hasta la 12162 y desde la 18971 hasta la 19028) Pertenece al 10 de marzo de 2013, 17 de marzo de 2013 y al 24 de marzo de 2013	NO	0

Tabla 29: Filtrado 1, vencimiento ESM13

Vencimiento ESU10

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	100093	NO	NO	NO	0
Proveedor2	100674	NO	NO	NO	0

Tabla 30: Filtrado 1, vencimiento ESU10

Vencimiento ESU11

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	99744	NO	NO	NO	0
Proveedor2	100548	NO	NO	NO	0

Tabla 31: Filtrado 1, vencimiento ESU11

Vencimiento ESU12

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	104833	NO	NO	NO	0
Proveedor2	105997	SI, a las 15:31 h	NO	NO	0

Tabla 32: Filtrado 1, vencimiento ESU12

Vencimiento ESU13

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	102003	NO	NO	NO	0
Proveedor2	102027	SI, a las 15:31 h	NO	NO	0

Tabla 33: Filtrado 1, vencimiento ESU13

Vencimiento ESZ10

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	97986	NO	NO	NO	0
Proveedor2	99285	NO	NO	NO	0

Tabla 34: Filtrado 1, vencimiento ESZ10

Vencimiento ESZ11

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	100344	NO	NO	NO	0
Proveedor2	101073	NO	SI, 59 líneas (desde la 54204 hasta la 54262) Pertenece al 30 de octubre de 2011	NO	0

Tabla 35: Filtrado 1, vencimiento ESZ11

Vencimiento ESM12

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	100241	NO	NO	NO	0
Proveedor2	100796	NO	SI, 58 líneas (desde la 48629 hasta la 48686) Pertenece al 28 de octubre de 2012	NO	0

Tabla 36: Filtrado 1, vencimiento ESZ12

Vencimiento ESZ13

	Número de líneas	¿Datos después del cierre, 15:30 h?	¿Hay datos algún sábado o domingo?	¿Volumen < 0?	Número de GAPs
Proveedor1	100441	NO	NO	NO	0
Proveedor2	100376	NO	SI, 61 líneas (en 13997, en 27648 y desde la 48139 hasta la 48197) Pertenece al 22 de septiembre de 2013, al 6 de octubre de 2013 y al 27 de octubre de 2013	NO	0

Tabla 37: Filtrado 1, vencimiento ESZ13

A continuación los resultados del Filtrado2:

Vencimiento ESH10

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 38: Filtrado 2, vencimiento ESH10

Vencimiento ESH11

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 39: Filtrado 2, vencimiento ESH11

Vencimiento ESH12

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 40: Filtrado 2, vencimiento ESH12

Vencimiento ESH13

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 41: Filtrado 2, vencimiento ESH13

Vencimiento ESM10

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 42: Filtrado 2, vencimiento ESM10

Vencimiento ESM11

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 43: Filtrado 2, vencimiento ESM11

Vencimiento ESM12

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 44: Filtrado 2, vencimiento ESM12

Vencimiento ESM13

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 45: Filtrado 2, vencimiento ESM13

Vencimiento ESU10

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 46: Filtrado 2, vencimiento ESU10

Vencimiento ESU11

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 47: Filtrado 2, vencimiento ESU11

Vencimiento ESU12

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 48: Filtrado 2, vencimiento ESU12

Vencimiento ESU13

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 49: Filtrado 2, vencimiento ESU13

Vencimiento ESZ10

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 50: Filtrado 2, vencimiento ESZ10

Vencimiento ESZ11

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 51: Filtrado 2, vencimiento ESZ11

Vencimiento ESZ12

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 52: Filtrado 2, vencimiento ESZ12

Vencimiento ESZ13

	¿Precio máximo, mínimo, cierre, apertura > 0?	¿Precio máximo ≥ precio mínimo, cierre, apertura?	¿Precio mínimo ≤ cierre, apertura?
Proveedor1	SI	SI	SI
Proveedor2	SI	SI	SI

Tabla 53: Filtrado 2, vencimiento ESZ13

A continuación los resultados del período horario de máxima negociación:

	Máximo volumen	Fecha (yyyyMMdd) máximo volumen	Hora (24h) máximo volumen	Hora (máximo volumen – 300) [24h]
ESH10	214712	20100113	18:04	13:04
ESH11	49414	20110228	22:00	17:00
ESH12	36749	20120229	22:01	17:01
ESH13	40015	20130102	22:00	17:00
ESM10	46966	20100331	22:00	17:00
ESM11	41644	20110531	22:00	17:00
ESM12	40456	20120531	22:00	17:00
ESM13	51842	20130531	22:01	17:01
ESU10	54222	20100831	22:00	17:01
ESU11	64190	20110831	22:00	17:00
ESU12	69548	20120629	22:00	17:00
ESU13	62380	20130628	22:00	17:00
ESZ10	43692	20101130	22:01	17:01
ESZ11	53306	20111130	22:00	17:00
ESZ12	40687	20120928	22:00	17:00
ESZ13	62222	20130930	22:00	17:00

Tabla 54: Máxima negociación – proveedor1

	Máximo volumen	Fecha (yyyyMMdd) máximo volumen	Hora (24h) máximo volumen	Hora (máximo volumen – 300) [24h]
ESH10	228784	20100113	18:04	13:04
ESH11	80343	20110228	22:00	17:00
ESH12	95900	20120229	22:01	17:01
ESH13	84624	20130228	22:00	17:00
ESM10	95155	20100430	22:00	17:00
ESM11	84923	20110331	22:00	17:00
ESM12	85047	20120531	22:00	17:00
ESM13	109250	20130531	22:00	17:00
ESU10	111901	20100831	22:00	17:00
ESU11	112305	20110831	22:00	17:00
ESU12	111769	20120629	22:00	17:00
ESU13	147482	20130628	22:00	17:00
ESZ10	105838	20101130	22:00	17:00
ESZ11	122516	20111031	21:00	16:00
ESZ12	96399	20120928	22:00	17:00
ESZ13	101695	20130930	22:00	17:00

Tabla 55: Máxima negociación – proveedor2

Se habrá podido dar cuenta el lector que ningún dato de volumen coincide. El dato del volumen muestra los paquetes de *trades* negociados y por lo general según el coste de la cuenta marcará la calidad del volumen (a pesar de no obtener mayor profundidad de mercado, tan sólo hablamos de paquetes y no qué lleva cada paquete). Si supiéramos el contenido del paquete, el coste de la cuenta se incrementaría considerablemente.

Para obtener una horquilla horaria, he considerado tomar en cuenta la hora de máxima negociación menos trescientos paquetes. Después de obtener las medidas, el intervalo horario de máxima negociación será desde las 17:00 horas hasta las 22:00 horas. Es decir, compararemos el volumen medio de negociación en esta horquilla cada día del presente vencimiento con el del siguiente vencimiento. Así conseguiremos saber cuándo aplicamos el *rollover*.

Llegados a este punto, pasaremos a construir el histórico continuo del futuro E-mini S&P 500, éste comenzará el 4 de enero de 2010 a las 00:01 horas (el 1 de enero es festivo) y terminará el viernes 20 de diciembre de 2013 a las 15:30 horas. A la vista de los resultados, el proveedor1 nos da mucha mayor seguridad debido a que ha pasado las pruebas de filtrado de manera más satisfactoria (no aparecen datos los sábados y domingos, también cierra a la hora de vencimiento). A pesar de ello, hemos borrado y estimado por mínimos cuadrados las irregularidades encontradas y reflejadas en las anteriores tablas. Para construir el gráfico continuo utilizaremos como base los vencimientos del proveedor1 ya filtrado.

A continuación la construcción del histórico continuo:

Tomando en cuenta el período ya comentado desde las 17:00 hasta las 22:00 horas.

	ESH10	ESM10
Volumen medio 1 marzo	1176.53488	9.05223881
Volumen medio 2 marzo	1278.2392	8.03333333
Volumen medio 3 marzo	1337.4485	8.19897959
Volumen medio 4 marzo	1099.48837	12.6527778
Volumen medio 5 marzo	1538.05316	11.8683128
Volumen medio 8 marzo	915.936877	18.4078212
Volumen medio 9 marzo	1672.51163	39.4542125
Volumen medio 10 marzo	1974.79402	107.730897
Volumen medio 11 marzo	733.478405	795.219269
Volumen medio 12 marzo	201.69103	1095.43854
Volumen medio 15 marzo	155.046512	1383.12957
Volumen medio 16 marzo	149.714286	1755.21595
Volumen medio 17 marzo	135.146667	1883.05316
Volumen medio 18 marzo	94.0604027	1659.61462
Día para el <i>rollover</i>	Jueves 11 de marzo de 2010	
Número de días adelantados al tercer viernes de vencimiento	6	
Precio cierre el 10 de marzo de 2010	1145.25	1135.25
Precio cierre el 11 de marzo de 2010	1150.25	1140.5
Split	$RAD = 1145.25 * \left(1 + \frac{1140.5 - 1145.25}{1140.5}\right) \cong 1140.5$ <p>Tenemos 9.75 puntos ($1150.25 - 1140.5$) que cotiza ESM10 por debajo de ESH10 en el momento de hacer el <i>rollover</i> y 10 puntos antes del <i>rollover</i>. Vemos que del día 10 al 11 de marzo en los dos vencimientos sube, pero si unimos 1145.25 (ESH10) y 1140.5 (ESM10) tenemos una bajada de ESH10 a ESM10 en vez de una subida como tendría que ser. Por tanto bajamos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(9.75+10)/2$ puntos ($1145.25 \rightarrow 1135.5$).</p>	

Tabla 56: *rollover* entre ESH10 y ESM10

	ESM10	ESU10
Volumen medio 1 junio	2122.96678	12.5267857
Volumen medio 2 junio	2115.17276	9.32217573
Volumen medio 3 junio	168.3289	14.7740586
Volumen medio 4 junio	2577.07309	21.9742647
Volumen medio 7 junio	1897.74419	21.2696246
Volumen medio 8 junio	3275.80399	54.1414141
Volumen medio 9 junio	2602.69767	96.3023256
Volumen medio 10 junio	1316.27907	1257.45183
Volumen medio 11 junio	497.212625	1455.56811
Volumen medio 14 junio	366.152824	1822.88372
Volumen medio 15 junio	301.043189	2071.02326
Volumen medio 16 junio	142.936877	1799.4485
Volumen medio 17 junio	130.541528	2064.32558
Día para el <i>rollover</i>	Viernes 11 de junio de 2010	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 10 de junio de 2010	1085.25	1048.5
Precio cierre el 11 de junio de 2010	1089.5	1081.25
Split	$RAD = 1085.25 * \left(1 + \frac{1081.25 - 1085.25}{1081.25}\right) \cong 1081.25$ <p>Tenemos 8.25 puntos ($1089.5 - 1081.25$) que cotiza ESU10 por debajo de ESM10 en el momento de hacer el <i>rollover</i> y 36.75 puntos antes del <i>rollover</i>. Vemos que del día 10 al 11 de junio en los dos vencimientos sube, pero si unimos 1085.25 (ESM10) y 1081.25 (ESU10) tenemos una bajada de ESM10 a ESU10 en vez de una subida como tendría que ser. Por tanto bajamos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(8.25+36.75)/2$ puntos ($1085.25 \rightarrow 1062.75$).</p>	

Tabla 57: *rollover* entre ESM10 y ESU10

	ESU10	ESZ10
Volumen medio 1 septiembre	1606.90033	5.72527473
Volumen medio 2 septiembre	1236.18272	6.11842105
Volumen medio 3 septiembre	1320.49169	11.817757
Volumen medio 6 septiembre	93.7666667	1.17647059
Volumen medio 7 septiembre	1174.17608	14.9592593
Volumen medio 8 septiembre	1313.38538	27.6819788
Volumen medio 9 septiembre	698.07309	558.578073
Volumen medio 10 septiembre	264.730897	872.611296
Volumen medio 13 septiembre	224.498339	1281.27907
Volumen medio 14 septiembre	208.481728	1632.88704
Volumen medio 15 septiembre	164.9066667	1417.77741
Volumen medio 16 septiembre	111.562712	1170.37542
Día para el rollover	Viernes 10 de septiembre de 2010	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 9 de septiembre de 2010	1101.75	1096
Precio cierre el 10 de septiembre de 2010	1109.5	1097
Split	$RAD = 1101.75 * \left(1 + \frac{1097 - 1101.75}{1097}\right) \cong 1097$ <p>Tenemos 12.5 puntos que cotiza ESZ10 por debajo de ESU10 en el momento de hacer el rollover y 5.75 puntos antes del rollover. Bajamos todo el histórico que tenemos hasta ahora, antes del rollover, $(12.5+5.75)/2$ puntos $(1101.75 \rightarrow 1092.5)$.</p>	

Tabla 58: rollover entre ESU10 y ESZ10

	ESZ10	ESH11
Volumen medio 1 diciembre	1884.98339	11.4652406
Volumen medio 2 diciembre	1514.64784	9.01604278
Volumen medio 3 diciembre	1255.42193	16.7696335
Volumen medio 6 diciembre	1099.81728	6.66666667
Volumen medio 7 diciembre	204.0897	35.5402299
Volumen medio 8 diciembre	1517.25581	48.7526132
Volumen medio 9 diciembre	642.418605	588.880399
Volumen medio 10 diciembre	296.624585	938.810631
Volumen medio 13 diciembre	193.438538	1205.07641
Volumen medio 14 diciembre	142.162791	1353.48505
Volumen medio 15 diciembre	178.050167	1678.42193
Volumen medio 16 diciembre	91.2331081	1384.49834
Día para el rollover	Viernes 10 de diciembre de 2010	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 9 de diciembre de 2010	1232.5	1225
Precio cierre el 10 de diciembre de 2010	1241	1227.75
Split	$RAD = 1232.5 * \left(1 + \frac{1227.75 - 1232.5}{1227.75}\right) \cong 1227.75$ <p>Tenemos 13.25 puntos que cotiza ESH11 por debajo de ESZ10 en el momento de hacer el rollover y 7.5 puntos antes del rollover. Bajamos todo el histórico que tenemos hasta ahora, antes del rollover, $(13.25+7.5)/2$ puntos $(1232.5 \rightarrow 1222.25)$.</p>	

Tabla 59: rollover entre ESZ10 y ESH11

	ESH11	ESM11
Volumen medio 1 marzo	2681.70432	10.111588
Volumen medio 2 marzo	2203.39867	12.5688073
Volumen medio 3 marzo	1711.12625	10.7149123
Volumen medio 4 marzo	2564.25914	15.4481328
Volumen medio 7 marzo	2503.90698	39.0646388
Volumen medio 8 marzo	1581.93355	30.1166078
Volumen medio 9 marzo	1563.3887	44.3472222
Volumen medio 10 marzo	1471.75415	1145.83389
Volumen medio 11 marzo	417.252492	1391.38538
Volumen medio 14 marzo	248.362126	2179.06645
Volumen medio 15 marzo	296.119601	2506.7309
Volumen medio 16 marzo	351.700997	4399.13289
Volumen medio 17 marzo	119.58804	1965.10963
Día para el rollover	Viernes 11 de marzo de 2011	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 10 de marzo de 2011	1294.5	1311
Precio cierre el 11 de marzo de 2011	1306.5	1288.75
Split	$RAD = 1294.5 * \left(1 + \frac{1288.75 - 1294.5}{1288.75}\right) \cong 1288.75$ <p>Tenemos 17.75 puntos en el momento de hacer el <i>rollover</i> y 16.5 puntos antes del <i>rollover</i>. Subimos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(16.5+17.75)/2$ puntos ($1294.5 \rightarrow 1311.75$).</p>	

Tabla 60: rollover entre ESH11 y ESM11

	ESM11	ESU11
Volumen medio 1 junio	2166.5814	4.70305677
Volumen medio 2 junio	2097.05648	7.17582418
Volumen medio 3 junio	1831.26246	7.86160714
Volumen medio 6 junio	1926.54817	10.9535865
Volumen medio 7 junio	1597.23256	31.1300448
Volumen medio 8 junio	2145.93023	57.1293706
Volumen medio 9 junio	861.631229	455
Volumen medio 10 junio	840.166113	1826.50498
Volumen medio 13 junio	321.056478	2220.01661
Volumen medio 14 junio	143.976744	1673.93688
Volumen medio 15 junio	286.358804	3267.39203
Volumen medio 16 junio	121.036545	2610.65449
Día para el <i>rollover</i>	Viernes 10 de junio de 2011	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 9 de junio de 2011	1287.5	1269.75
Precio cierre el 10 de junio de 2011	1272.5	1282.25
Split	$RAD = 1287.5 * \left(1 + \frac{1282.25 - 1287.5}{1282.25}\right) \cong 1282.25$ <p>Tenemos 9.75 puntos en el momento de hacer el <i>rollover</i> y 17.75 puntos antes del <i>rollover</i>. Bajamos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(9.75+17.75)/2$ puntos ($1287.5 \rightarrow 1273.75$).</p>	

Tabla 61: *rollover* entre ESM11 y ESU11

	ESU11	ESZ11
Volumen medio 1 septiembre	2083.9701	6.34680135
Volumen medio 2 septiembre	1640.81728	4.95794393
Volumen medio 5 septiembre	469.3	7.20833333
Volumen medio 6 septiembre	2564.55482	25.2132867
Volumen medio 7 septiembre	2212.95681	48.3198653
Volumen medio 8 septiembre	1189.34884	883.847176
Volumen medio 9 septiembre	868.013289	2631.78405
Volumen medio 12 septiembre	394.269103	2838.78738
Volumen medio 13 septiembre	211.053156	2239.89037
Volumen medio 14 septiembre	302.820598	2679.0897
Volumen medio 15 septiembre	185.830565	2088.84718
Día para el <i>rollover</i>	Viernes 9 de septiembre de 2011	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 8 de septiembre de 2011	1186.5	1194.25
Precio cierre el 9 de septiembre de 2011	1158.25	1180.5
Split	$RAD = 1186.5 * \left(1 + \frac{1180.5 - 1186.5}{1180.5}\right) \cong 1180.5$ <p>Tenemos 22.25 puntos en el momento de hacer el <i>rollover</i> y 7.75 puntos antes del <i>rollover</i>. Subimos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(22.25+7.75)/2$ puntos ($1186.5 \rightarrow 1201.5$).</p>	

Tabla 62: *rollover* entre ESU11 y ESZ11

	ESZ11	ESH12
Volumen medio 1 diciembre	1747.83721	9.06024096
Volumen medio 2 diciembre	1654.33223	9.01724138
Volumen medio 5 diciembre	2128.2392	10.7894737
Volumen medio 6 diciembre	1817.7907	8.57345972
Volumen medio 7 diciembre	2354.25914	74.090604
Volumen medio 8 diciembre	1420.51163	784.48505
Volumen medio 9 diciembre	506.169435	1656.4485
Volumen medio 12 diciembre	262.784053	1808.2093
Volumen medio 13 diciembre	324.385382	2117.78738
Volumen medio 14 diciembre	134.920266	2172.48173
Volumen medio 15 diciembre	72.6179402	1557.43854
Día para el <i>rollover</i>	Viernes 9 de diciembre de 2011	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 8 de diciembre de 2011	1232.75	1259.75
Precio cierre el 9 de diciembre de 2011	1259.75	1228.5
Split	$RAD = 1232.75 * \left(1 + \frac{1228.5 - 1232.75}{1228.5}\right) \cong 1228.5$ <p>Tenemos 31.25 puntos en el momento de hacer el <i>rollover</i> y 27 puntos antes del <i>rollover</i>. Subimos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(31.25+27)/2$ puntos ($1232.75 \rightarrow 1261.75$).</p>	

Tabla 63: *rollover* entre ESZ11 y ESH12

	ESH12	ESM12
Volumen medio 1 marzo	1473.90365	32.1123596
Volumen medio 2 marzo	1283.90698	3.95209581
Volumen medio 5 marzo	1214.6412	9.26442308
Volumen medio 6 marzo	1807.12292	17.528777
Volumen medio 7 marzo	1533.2093	43.8108108
Volumen medio 8 marzo	995.531561	348.289037
Volumen medio 9 marzo	343.551495	912.996678
Volumen medio 12 marzo	301.916944	1033.21262
Volumen medio 13 marzo	314.295681	2234.36213
Volumen medio 14 marzo	136.142857	1507.45515
Volumen medio 15 marzo	110.322259	1373.33555
Día para el <i>rollover</i>	Viernes 9 de marzo de 2012	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 8 de marzo de 2012	1365.5	1346.5
Precio cierre el 9 de marzo de 2012	1372	1360
Split	$RAD = 1365.5 * \left(1 + \frac{1360 - 1365.5}{1360}\right) \cong 1360$ <p>Tenemos 12 puntos en el momento de hacer el <i>rollover</i> y 19 puntos antes del <i>rollover</i>. Bajamos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(19+12)/2$ puntos ($1365.5 \rightarrow 1350$).</p>	

Tabla 64: *rollover* entre ESH12 y ESM12

	ESM12	ESU12
Volumen medio 1 junio	2166.58804	15.8536585
Volumen medio 4 junio	1775.18272	19.3937008
Volumen medio 5 junio	1623.30565	32.0501672
Volumen medio 6 junio	2166.40532	56.9799331
Volumen medio 7 junio	1435.90033	453.252492
Volumen medio 8 junio	540.448505	1201.38206
Volumen medio 11 junio	293.578073	2088.52159
Volumen medio 12 junio	165.372093	1703.46512
Volumen medio 13 junio	114.27907	1904.64452
Volumen medio 14 junio	163.923333	2303.28239
Día para el <i>rollover</i>	Viernes 8 de junio de 2012	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 7 de junio de 2012	1317.5	1309.25
Precio cierre el 8 de junio de 2012	1317.5	1310
Split	$RAD = 1317.5 * \left(1 + \frac{1310 - 1317.5}{1310}\right) \cong 1310$ <p>Tenemos 7.5 puntos en el momento de hacer el <i>rollover</i> y 8.25 puntos antes del <i>rollover</i>. Bajamos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(8.25+7.75)/2$ puntos ($1317.5 \rightarrow 1309.5$).</p>	

Tabla 65: *rollover* entre ESM12 y ESU12

	ESU12	ESZ12
Volumen medio 3 septiembre	228.533333	5
Volumen medio 4 septiembre	1228.89037	4.65671642
Volumen medio 5 septiembre	815.083056	2.65644172
Volumen medio 6 septiembre	1328.74086	9.17509728
Volumen medio 7 septiembre	795.448505	10.2398374
Volumen medio 10 septiembre	795.946844	17.0987654
Volumen medio 11 septiembre	869.348837	19.7023411
Volumen medio 12 septiembre	1014.71429	48.7458194
Volumen medio 13 septiembre	2643.09635	1079.48173
Volumen medio 14 septiembre	484.724252	1441
Volumen medio 17 septiembre	200.269103	1060.66777
Volumen medio 18 septiembre	109.023569	974.647841
Volumen medio 19 septiembre	80.2020548	959.664452
Volumen medio 20 septiembre	78.5752508	1231.35548
Día para el rollover	Viernes 14 de septiembre de 2012	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 13 de septiembre de 2012	1457.25	1431.5
Precio cierre el 14 de septiembre de 2012	1457.25	1451.25
Split	$RAD = 1457.25 * \left(1 + \frac{1451.25 - 1457.25}{1451.25}\right) \cong 1288.75$ <p>Tenemos 6 puntos en el momento de hacer el rollover y 25.75 puntos antes del rollover. Bajamos todo el histórico que tenemos hasta ahora, antes del rollover, $(6+25.75)/2$ puntos ($1457.25 \rightarrow 1441.25$).</p>	

Tabla 66: rollover entre ESU12 y ESZ12

	ESZ12	ESH13
Volumen medio 3 diciembre	1302.20598	3.14285714
Volumen medio 4 diciembre	1409.87375	7.59210526
Volumen medio 5 diciembre	2023.98671	7.90376569
Volumen medio 6 diciembre	1035.02658	10.5933014
Volumen medio 7 diciembre	1312.53156	13.2452107
Volumen medio 10 diciembre	767.74751	13.8226415
Volumen medio 11 diciembre	1527.20598	48.6315789
Volumen medio 12 diciembre	1855.95681	110.95302
Volumen medio 13 diciembre	1449.88372	442.704319
Volumen medio 14 diciembre	276.541528	839.398671
Volumen medio 17 diciembre	227.043189	1543.98007
Volumen medio 18 diciembre	204.212625	1692.66113
Volumen medio 19 diciembre	102.813953	1478.94684
Volumen medio 20 diciembre	112.356667	1240.18937
Día para el <i>rollover</i>	Viernes 14 de diciembre de 2012	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 13 de diciembre de 2012	1419.75	1417.5
Precio cierre el 14 de diciembre de 2012	1411.25	1416.5
Split	$RAD = 1419.75 * \left(1 + \frac{1416.5 - 1419.75}{1416.5}\right) \cong 1416.5$ <p>Tenemos 5.25 puntos en el momento de hacer el <i>rollover</i> y 2.25 puntos antes del <i>rollover</i>. Bajamos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(5.25+2.25)/2$ puntos ($1419.75 \rightarrow 1416$).</p>	

Tabla 67: *rollover* entre ESZ12 y ESH13

	ESH13	ESM13
Volumen medio 1 marzo	1881.24252	9.87148594
Volumen medio 4 marzo	1698.93023	25.4273859
Volumen medio 5 marzo	1534.3289	41.3287671
Volumen medio 6 marzo	1096.11628	36.510274
Volumen medio 7 marzo	733.72093	216.528239
Volumen medio 8 marzo	385.66113	951.750831
Volumen medio 11 marzo	215.903654	1086.63455
Volumen medio 12 marzo	177.784053	1444.69435
Volumen medio 13 marzo	115.747508	919.33887
Volumen medio 14 marzo	121.385382	989.328904
Día para el <i>rollover</i>	Viernes 8 de marzo de 2013	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 7 de marzo de 2013	1544	1534.75
Precio cierre el 8 de marzo de 2013	1550.25	1537.75
Split	$RAD = 1544 * \left(1 + \frac{1537.75 - 1544}{1537.75}\right) \cong 1537.75$ <p>Tenemos 12.5 puntos en el momento de hacer el <i>rollover</i> y 9.25 puntos antes del <i>rollover</i>. Bajamos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(12.5+9.25)/2$ puntos ($1544 \rightarrow 1533.25$).</p>	

Tabla 68: *rollover* entre ESH13 y ESM13

	ESM13	ESU13
Volumen medio 3 junio	1943.36545	18.9926471
Volumen medio 4 junio	2054.70432	14.1607143
Volumen medio 5 junio	2490.78738	15.1188811
Volumen medio 6 junio	2708.19269	28.1517241
Volumen medio 7 junio	1641.56478	22.9421488
Volumen medio 10 junio	1293.70432	17.9601449
Volumen medio 11 junio	2000.27907	32.6013289
Volumen medio 12 junio	1941.15947	84.3388704
Volumen medio 13 junio	1628.88704	513.740864
Volumen medio 14 junio	409.827243	1235.48505
Volumen medio 17 junio	257.259136	1379.12625
Volumen medio 18 junio	165.272425	1130.2093
Volumen medio 19 junio	209.829431	2174.57807
Volumen medio 20 junio	154.637874	2735.33223
Día para el rollover	Viernes 14 de junio de 2013	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 13 de junio de 2013	1637.5	1605.5
Precio cierre el 14 de junio de 2013	1623.5	1632.25
Split	$RAD = 1637.5 * \left(1 + \frac{1632.25 - 1637.5}{1632.25}\right) \cong 1632.25$ <p>Tenemos 8.75 puntos en el momento de hacer el rollover y 32 puntos antes del rollover. Bajamos todo el histórico que tenemos hasta ahora, antes del rollover, $(32+8.75)/2$ puntos ($1637.5 \rightarrow 1617.25$).</p>	

Tabla 69: rollover entre ESM13 y ESU13

	ESU13	ESZ13
Volumen medio 2 septiembre	182.8	0.77777778
Volumen medio 3 septiembre	2051.51163	6.17479675
Volumen medio 4 septiembre	1509.87375	7.98062016
Volumen medio 5 septiembre	924.860465	6.08658009
Volumen medio 6 septiembre	1638.02326	12.9925651
Volumen medio 9 septiembre	1383.03322	18.1672727
Volumen medio 10 septiembre	1505.38206	57.3255814
Volumen medio 11 septiembre	1536.44186	55.2126246
Volumen medio 12 septiembre	878.946844	282.770764
Volumen medio 13 septiembre	229.332226	481.026578
Volumen medio 16 septiembre	243.445183	996.245847
Volumen medio 17 septiembre	113.533557	580.232558
Volumen medio 18 septiembre	312.38206	2364.58804
Volumen medio 19 septiembre	85.2524917	1112.39203
Día para el <i>rollover</i>	Viernes 13 de septiembre de 2013	
Número de días adelantados al tercer viernes de vencimiento	5	
Precio cierre el 12 de septiembre de 2013	1685.5	1683
Precio cierre el 13 de septiembre de 2013	1689.5	1678
Split	$RAD = 1685.5 * \left(1 + \frac{1678 - 1685.5}{1678}\right) \cong 1678$ <p>Tenemos 11.5 puntos en el momento de hacer el <i>rollover</i> y 2.5 puntos antes del <i>rollover</i>. Bajamos todo el histórico que tenemos hasta ahora, antes del <i>rollover</i>, $(11.5+2.5)/2$ puntos ($1685.5 \rightarrow 1678.5$).</p>	

Tabla 70: *rollover* entre ESU13 y ESZ13

Comprobamos que la tendencia para realizar el rollover del futuro E-mini S&P 500 es justamente una semana antes al día marcado oficialmente (tercer viernes del mes). Antes de continuar con los cálculos para hallar el intervalo de confianza para el valor de la pendiente en un tramo lateral, pasemos a ver unas representaciones gráficas de cómo ha quedado el histórico continuo filtrado con el *rollover* y el *split* aplicado. También la muestra IS y la muestra OOS por separado.

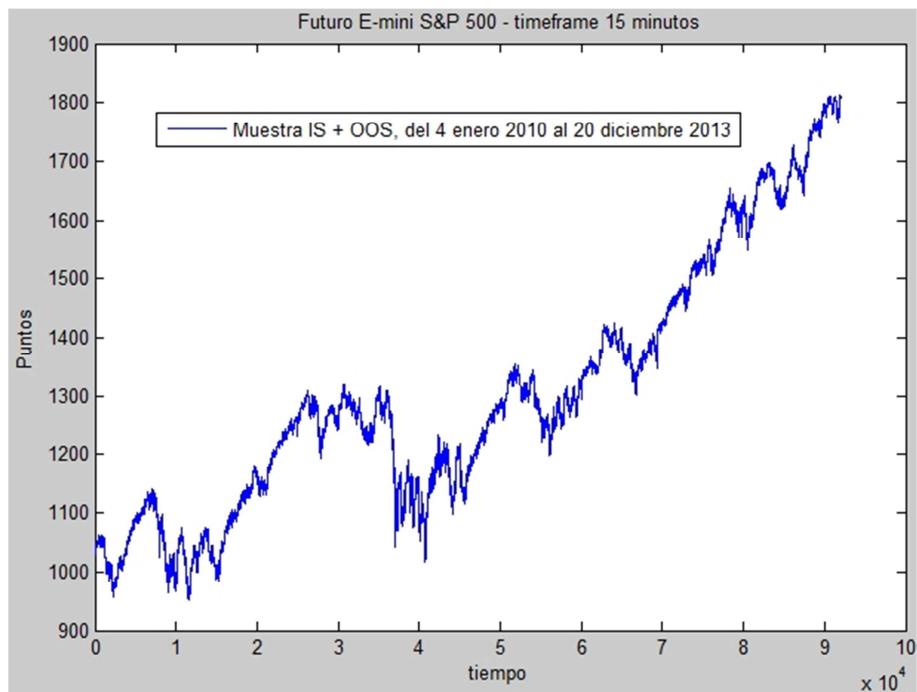


Figura 94: Muestra IS+OOS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

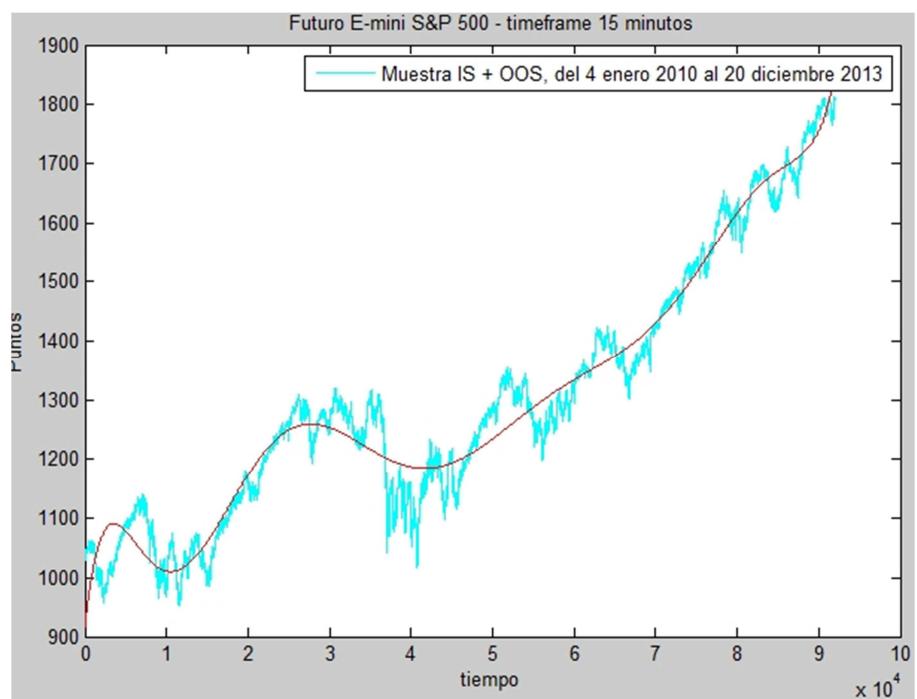


Figura 95: Muestra IS + OOS, ajuste – precio cierre – E-mini S&P 500 - 15 min

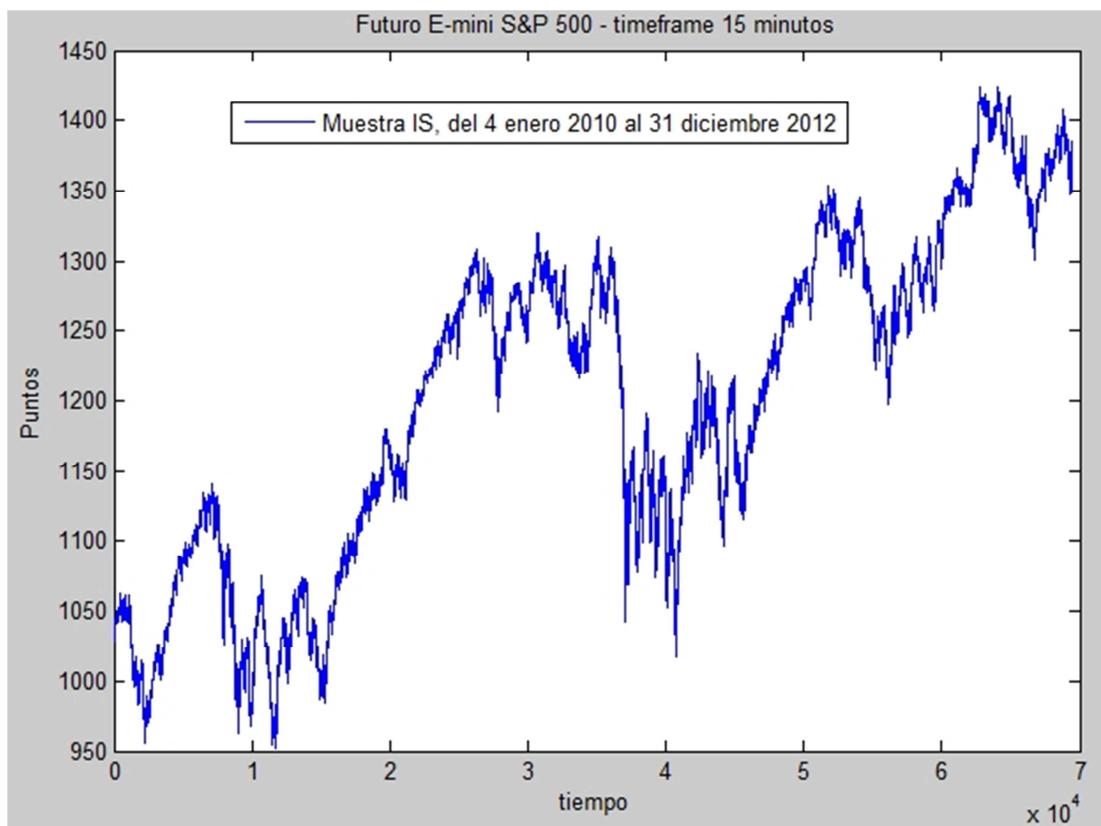


Figura 96: Muestra IS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

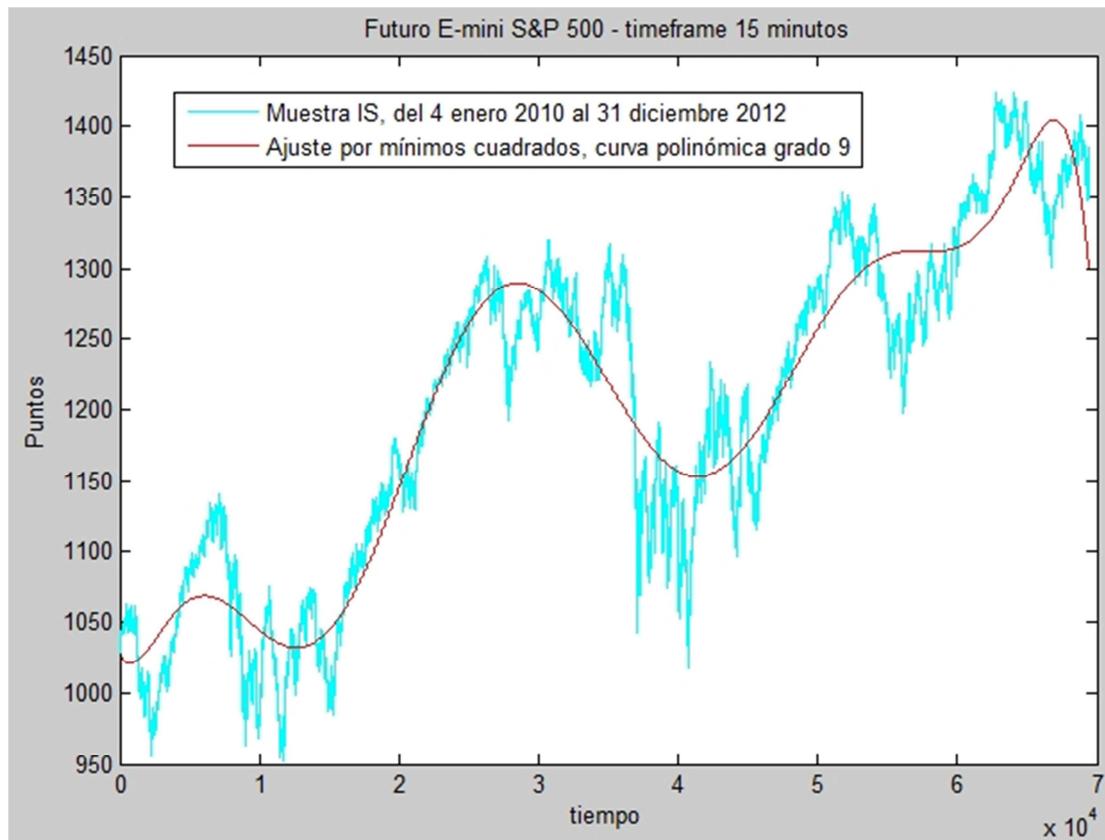


Figura 97: Muestra IS, ajuste – precio cierre – E-mini S&P 500 - 15 min

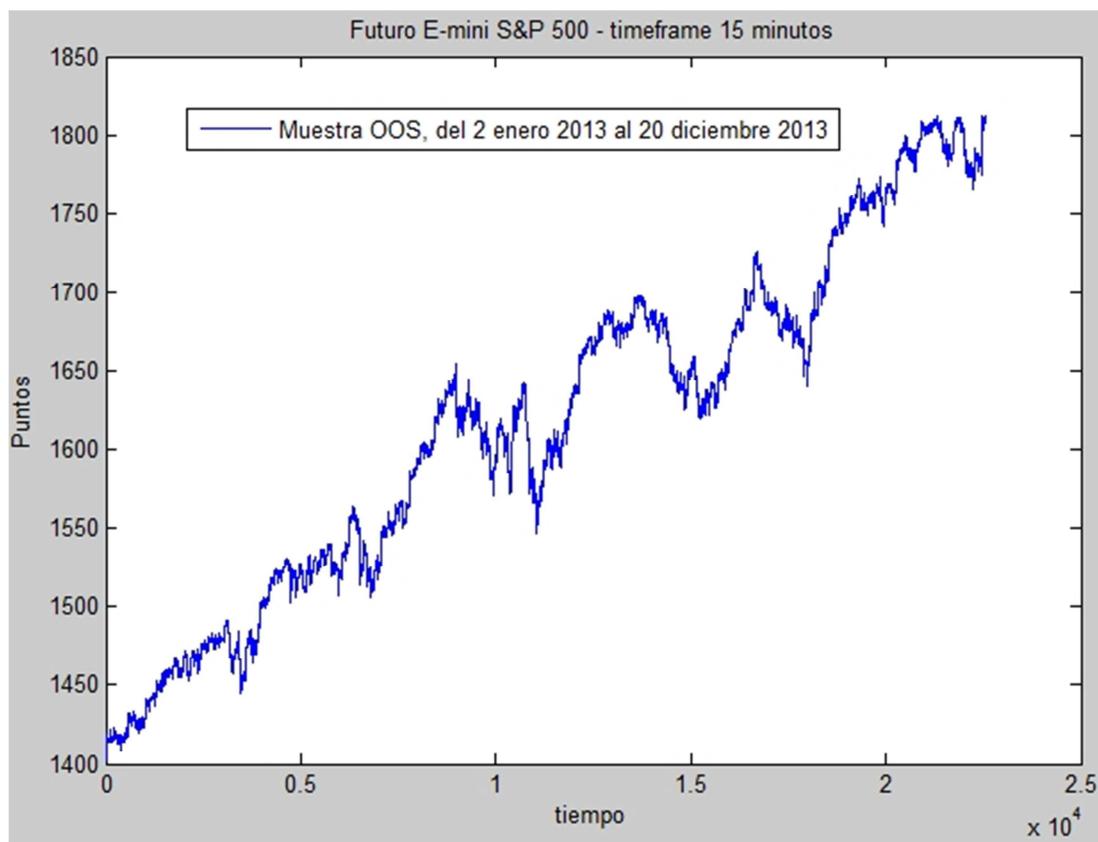


Figura 98: Muestra OOS, histórico continuo – precio cierre – E-mini S&P 500 - 15 min

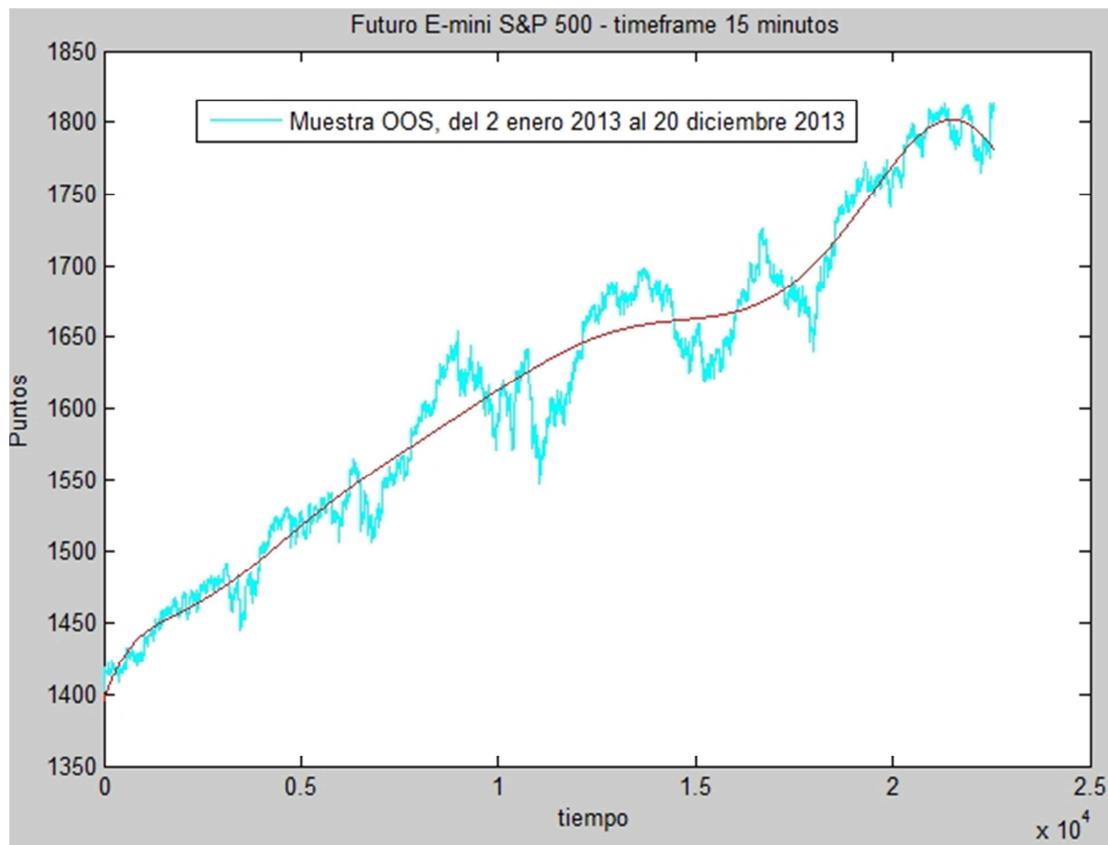


Figura 99: Muestra OOS, ajuste – precio cierre – E-mini S&P 500 - 15 min

Finalmente procedemos a marcar sobre el gráfico los tramos laterales, será un período de los máximos y mínimos registrados en la curva. Fuera de dichos tramos, la pendiente, de la recta tangente a la curva ajustada, muestre tendencia bajista o alcista. Recordamos que se trata de un procedimiento mejorable, todavía me queda por estudiar nuevos procedimientos, debido a que se trata de un concepto novedoso el hecho de preprocesar los datos sobreajustando con una curva para obtener los resultados de optimización divididos en tendencia alcista | bajista | lateral. Gracias a los citados resultados ya divididos podremos tomar decisiones del tipo “Mi sistema se comporta mejor en tramo lateral” y así elaborar por ejemplo un portfolio de sistemas más eficiente.

Periodo 1

	Fecha y hora comienzo yyyyMMdd HHmm	Fecha y hora comienzo yyyyMMdd HHmm	Posición inicial	Posición final
Tramo lateral	20100104 1900	20100121 0726	437	1097

Figura 71: Período 1 de un tramo lateral

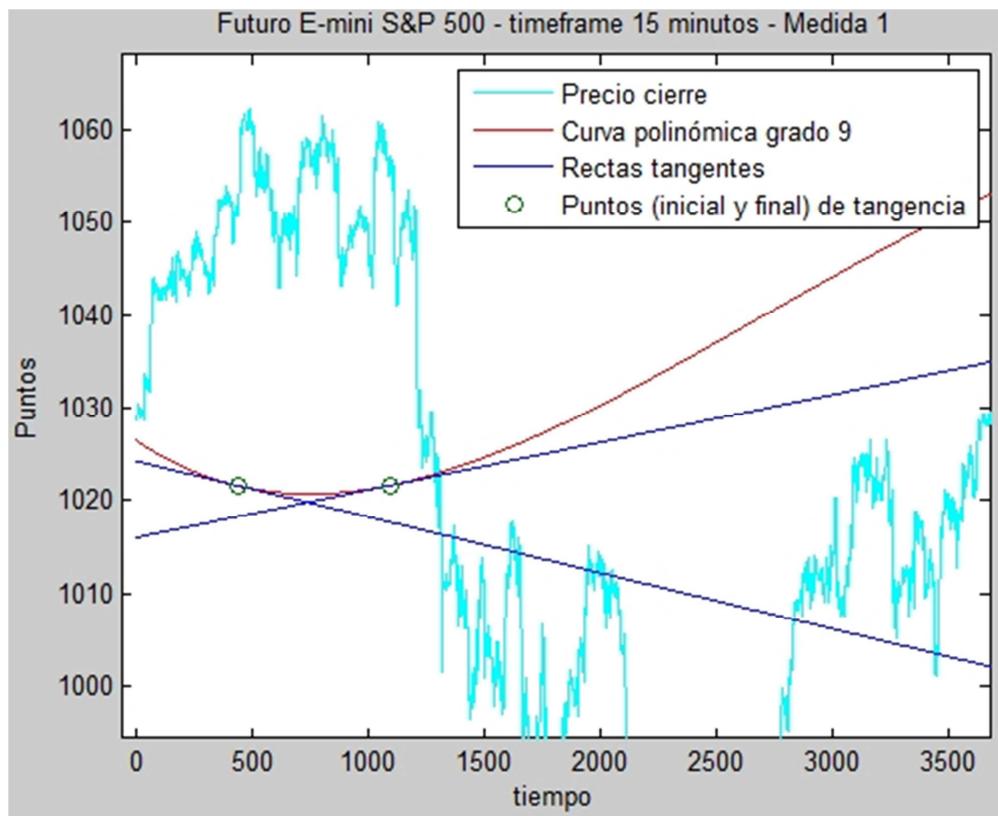


Figura 100: Período 1 de un tramo lateral

Periodo 2

	Fecha y hora comienzo yyyyMMdd HHmm	Fecha y hora comienzo yyyyMMdd HHmm	Posición inicial	Posición final
Tramo lateral	20100412 2134	20100430 0633	6250	7450

Figura 72: Período 2 de un tramo lateral

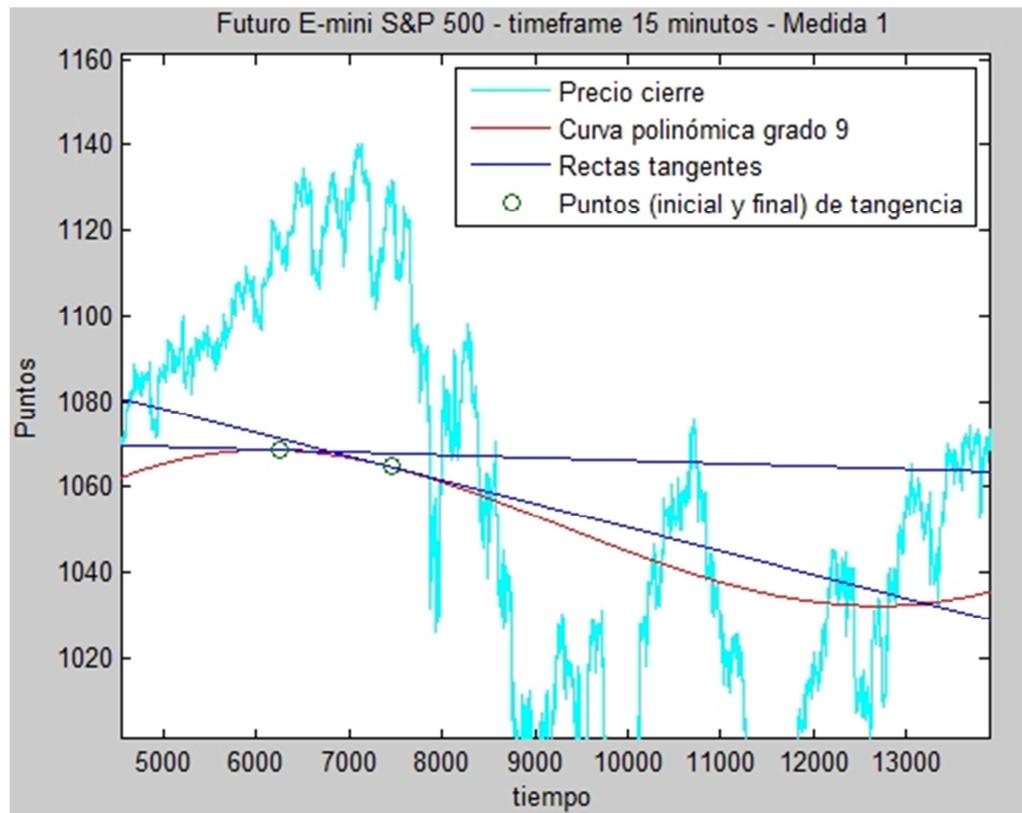


Figura 101: Período 2 de un tramo lateral

Periodo 3

	Fecha y hora comienzo yyyyMMdd HHmm	Fecha y hora comienzo yyyyMMdd HHmm	Posición inicial	Posición final
Tramo lateral	20100715 1420	20100818 0308	12340	14460

Figura 73: Período 3 de un tramo lateral

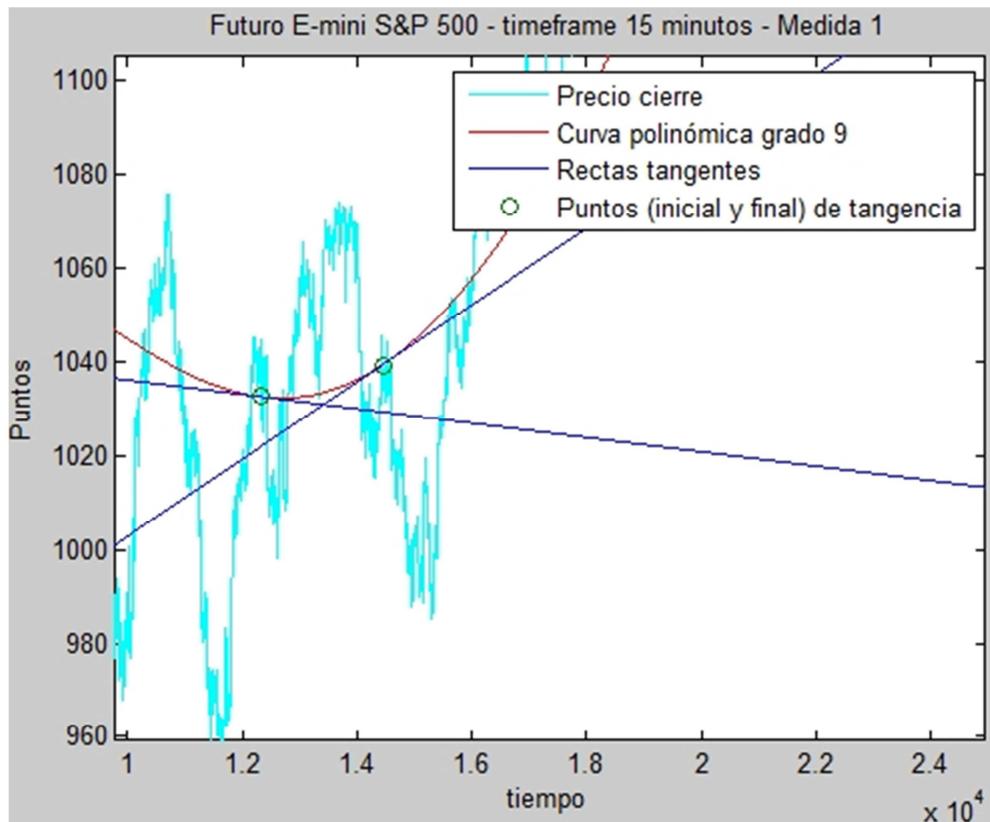


Figura 102: Período 3 de un tramo lateral

Periodo 4

	Fecha y hora comienzo yyyyMMdd HHmm	Fecha y hora comienzo yyyyMMdd HHmm	Posición inicial	Posición final
Tramo lateral	20110228 0403	20110309 2006	26710	27410

Tabla 74: Período 4 de un tramo lateral

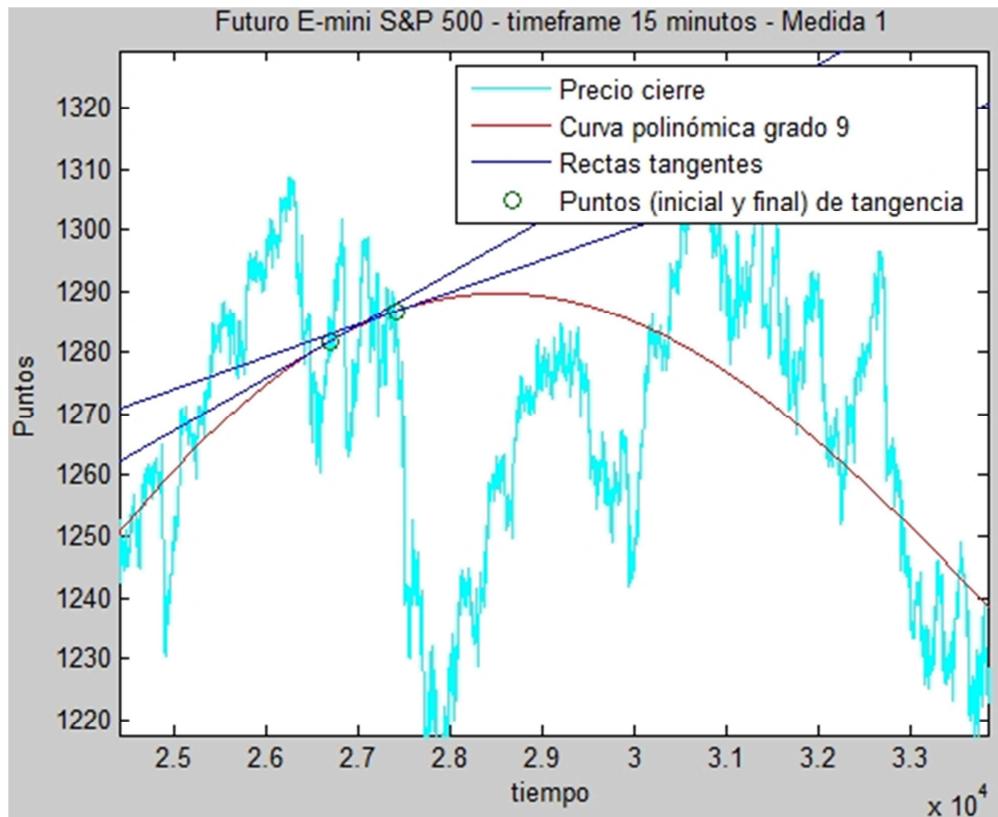


Figura 103: Período 4 de un tramo lateral

Periodo 5

	Fecha y hora comienzo yyyyMMdd HHmm	Fecha y hora comienzo yyyyMMdd HHmm	Posición inicial	Posición final
Tramo lateral	20110331 0317	20110418 1717	28800	29440

Tabla 75: Período 5 de un tramo lateral

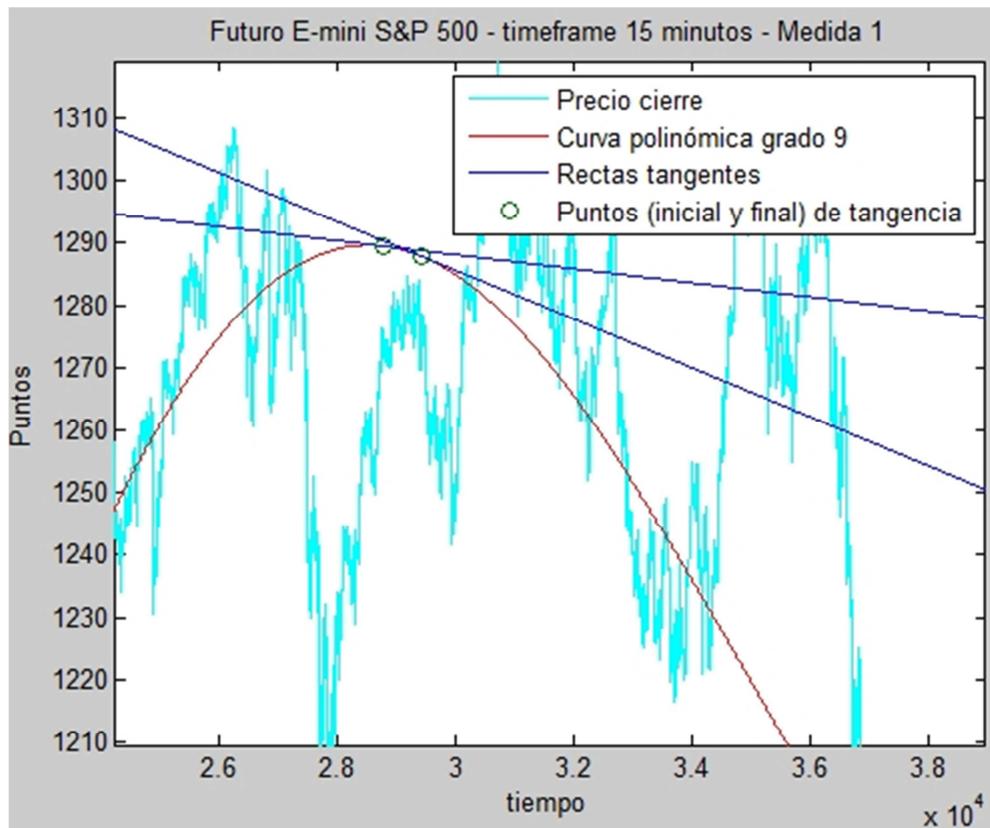


Figura 104: Período 5 de un tramo lateral

Periodo 6

	Fecha y hora comienzo yyyyMMdd HHmm	Fecha y hora comienzo yyyyMMdd HHmm	Posición inicial	Posición final
Tramo lateral	20110815 2014	20110927 1127	37510	40280

Tabla 76: Período 6 de un tramo lateral

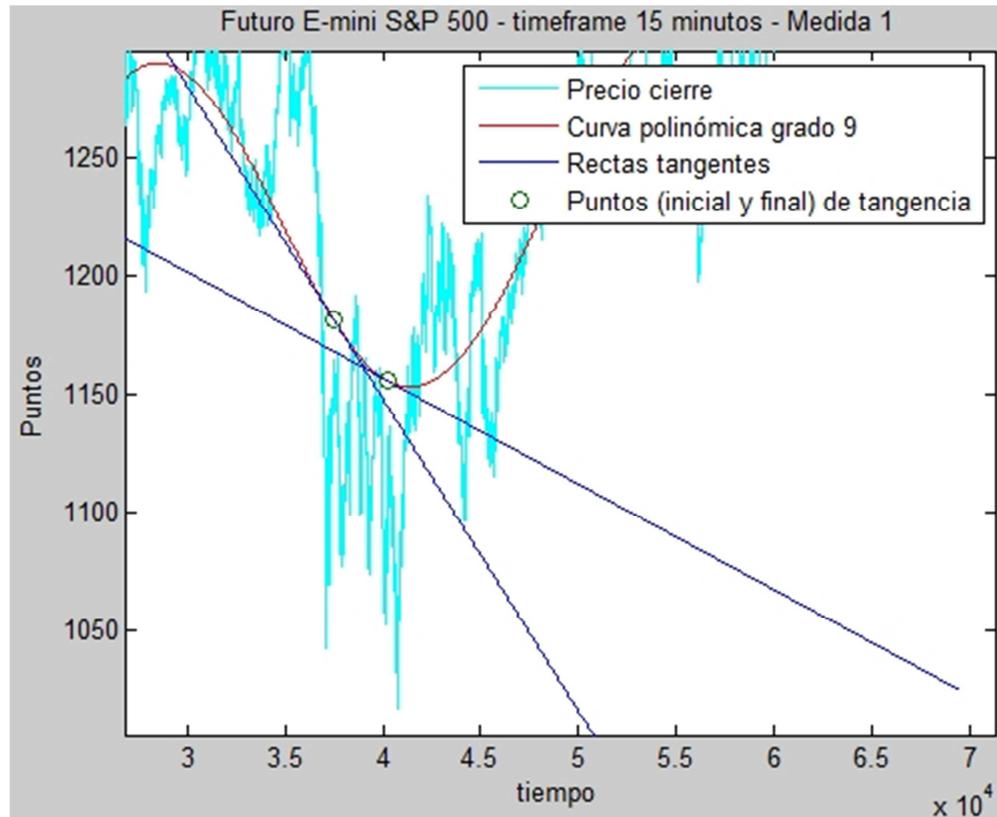


Figura 105: Período 6 de un tramo lateral

Periodo 7

	Fecha y hora comienzo yyyyMMdd HHmm	Fecha y hora comienzo yyyyMMdd HHmm	Posición inicial	Posición final
Tramo lateral	20111031 0329	20111115 1442	42440	43490

Tabla 77: Período 7 de un tramo lateral

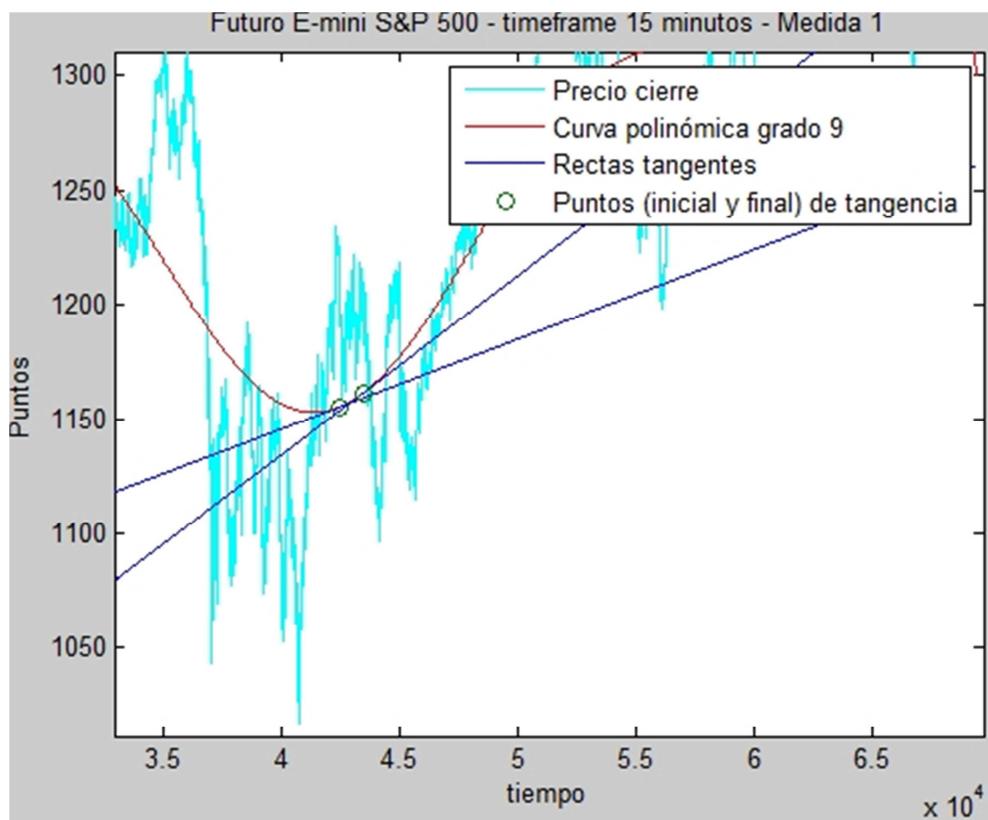


Figura 106: Período 7 de un tramo lateral

Periodo 8

	Fecha y hora comienzo yyyyMMdd HHmm	Fecha y hora comienzo yyyyMMdd HHmm	Posición inicial	Posición final
Tramo lateral	20120730 0608	20120802 1800	59680	60000

Tabla 78: Período 8 de un tramo lateral

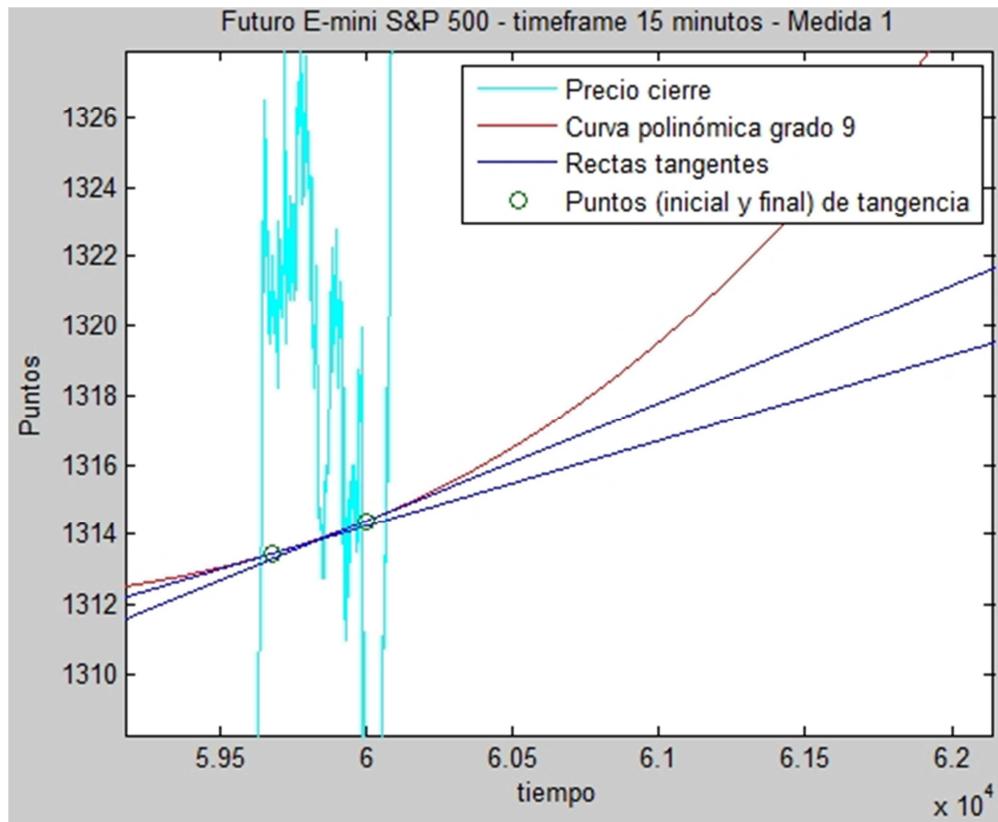


Figura 107: Período 8 de un tramo lateral

Periodo 9

	Fecha y hora comienzo yyyyMMdd HHmm	Fecha y hora comienzo yyyyMMdd HHmm	Posición inicial	Posición final
Tramo lateral	20121024 0759	20121106 0854	65250	66010

Tabla 79: Período 9 de un tramo lateral

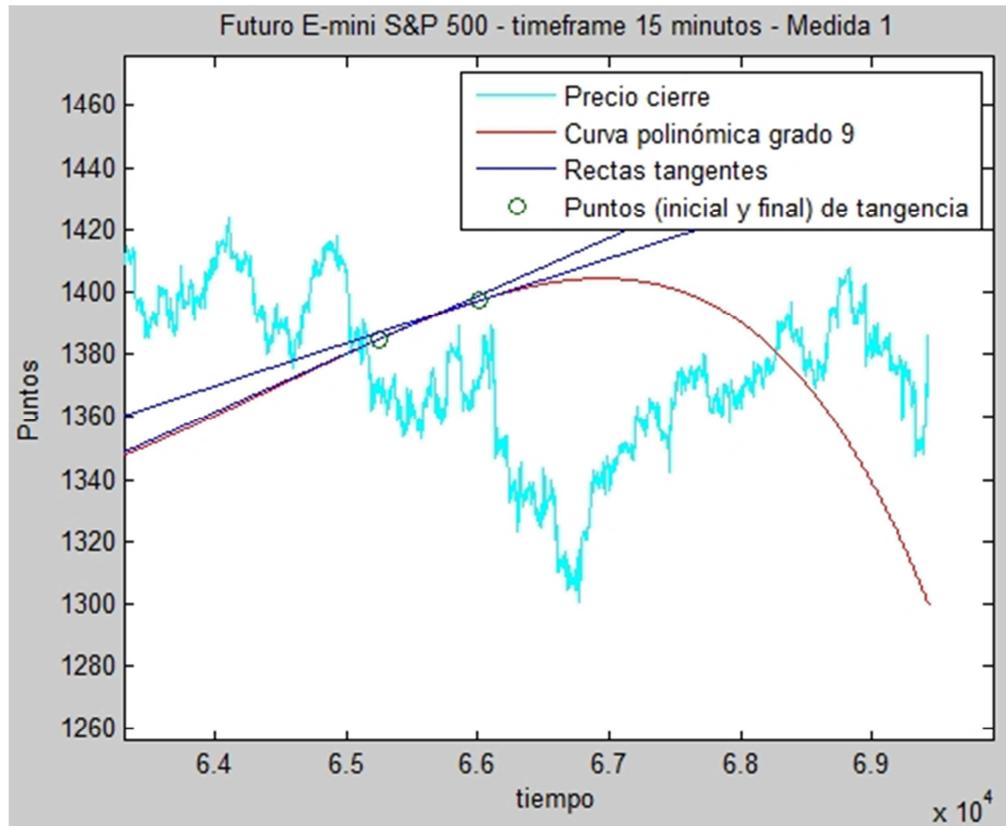


Figura 108: Período 9 de un tramo lateral

Periodo 10

	Fecha y hora comienzo yyyyMMdd HHmm	Fecha y hora comienzo yyyyMMdd HHmm	Posición inicial	Posición final
Tramo lateral	20121224 0759	20121204 1503	67820	68190

Tabla 80: Período 10 de un tramo lateral

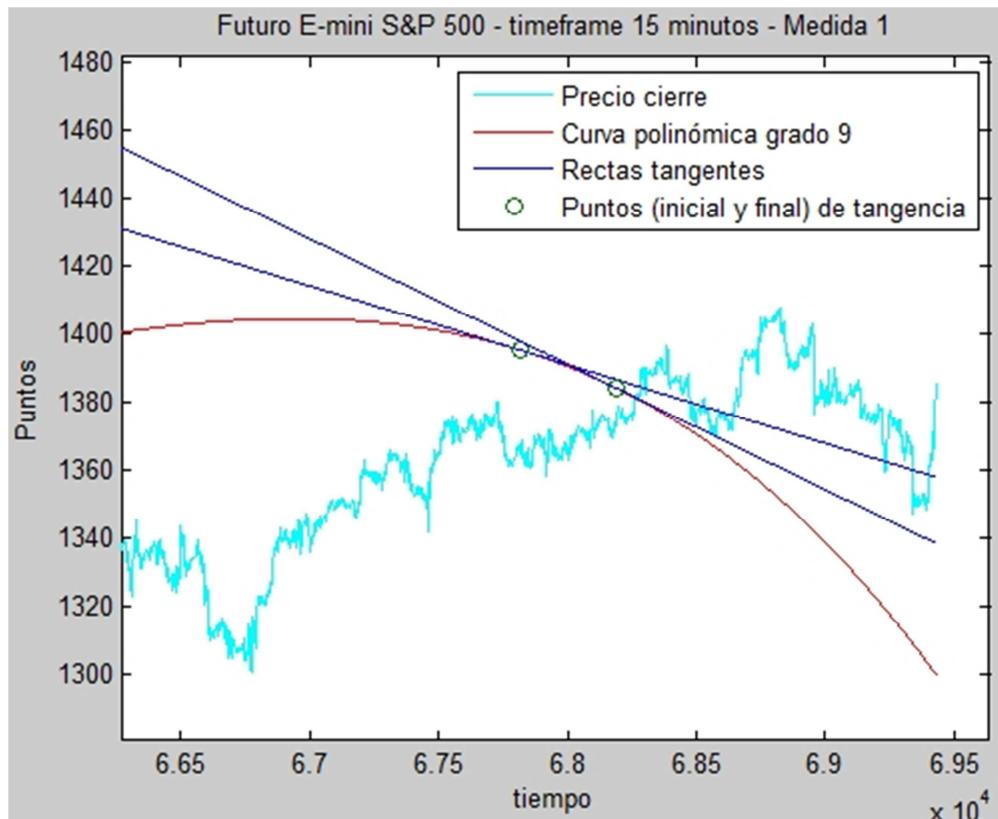


Figura 109: Período 10 de un tramo lateral

ANEXO III: Informe detallado - resultados fase 2

El procedimiento de presentación de resultados será; nombre experimento, muestra de combinaciones seleccionada, función *fitness* aplicada, filtro utilizado, tiempo de ejecución de la optimización, carga media del sistema y consumo *RAM*, PBO obtenido, nube de puntos del valor de la función *fitness* IS, nube de puntos del valor de la función *fitness* OOS, histograma de los valores de la función *fitness* IS, mejor combinación de parámetros (mejor *equity* IS), retorno - máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS, número de trades en tramos alcistas | bajistas | laterales, gráfica de la *equity* IS superpuesto con la *equity* OOS de la mejor combinación de parámetros.

Es importante tomar en cuenta que no hay diferencias, entre los experimentos, a razón el filtrado mínimo por número de contratos debido a que ha dado la casualidad que el presente sistema en la muestra de datos (IS) ejecutada realiza más de 300 operaciones (*trades*) siempre.

Para tener una referencia computacional en los resultados obtenidos, mostramos el mapa de la *RAM* del sistema nada más arrancar:



Figura 110: Consumo de *RAM* inicial – Sistema

Y una vez arrancada la aplicación MATLAB:



Figura 111: Consumo de *RAM* inicial con MATLAB arrancado – Sistema

Nombre del experimento: Experimento 1,2 y 3

Número de combinaciones: 1764

Función *fitness*: SQN

Filtro: 50, 100 y 300

Tiempo de ejecución de la optimización (paralelo) = 2.76 minutos

Tiempo de ejecución de la optimización (no paralelo) = 3.09 minutos



Figura 112: Consumo *RAM* – Experimento 1, 2 y 3

Como mínimo necesitamos una variable que ocupa $69441 * 1764 * 8 = 979.95$ MB

1	[1.9%]	4	[0.0%]	7	[100.0%]	10	[100.0%]
2	[100.0%]	5	[100.0%]	8	[0.5%]	11	[0.0%]
3	[1.4%]	6	[3.7%]	9	[100.0%]	12	[100.0%]

Figura 113: Carga del sistema – 6 cores – Experimento 1, 2 y 3

Load average: 2.51 3.58 4.01

Figura 114: Carga media del sistema – Experimento 1, 2 y 3

PBO = 100/252 (39.68 %)

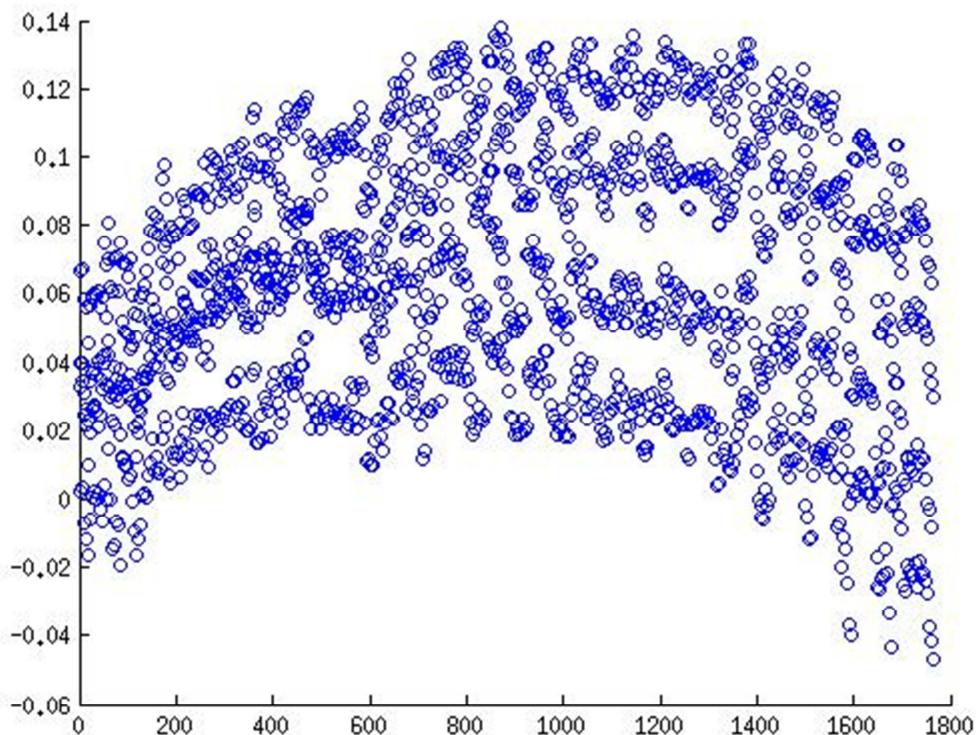


Figura 115: Nube de puntos del resultado de la función *fitness IS* – Experimento 1, 2 y 3

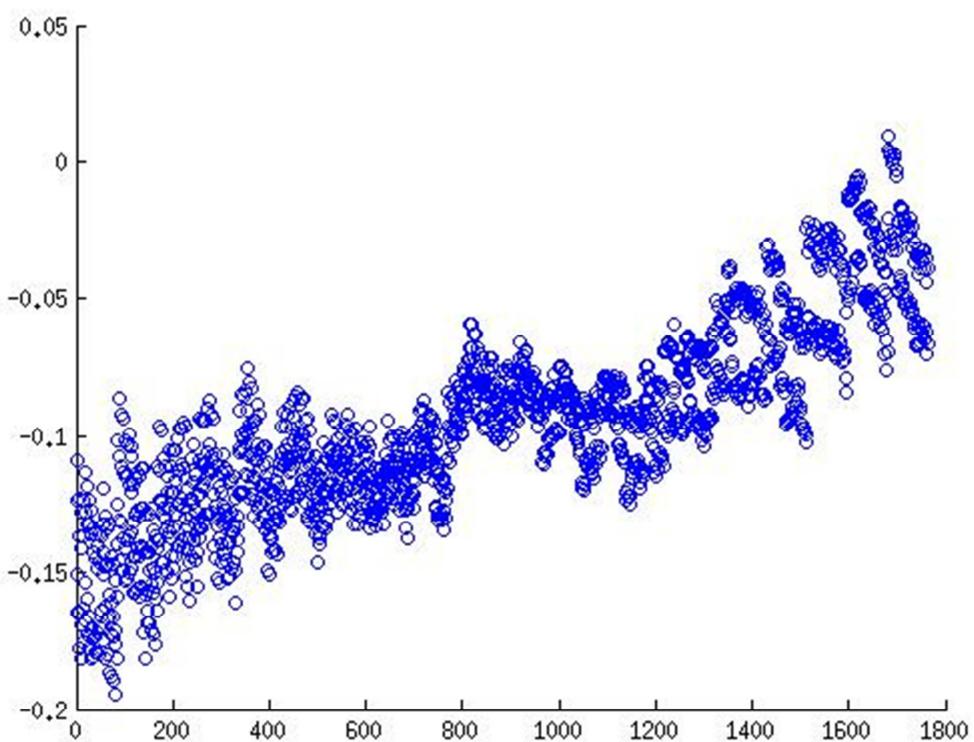


Figura 116: Nube de puntos del resultado de la función *fitness* OOS – Experimento 1, 2 y 3

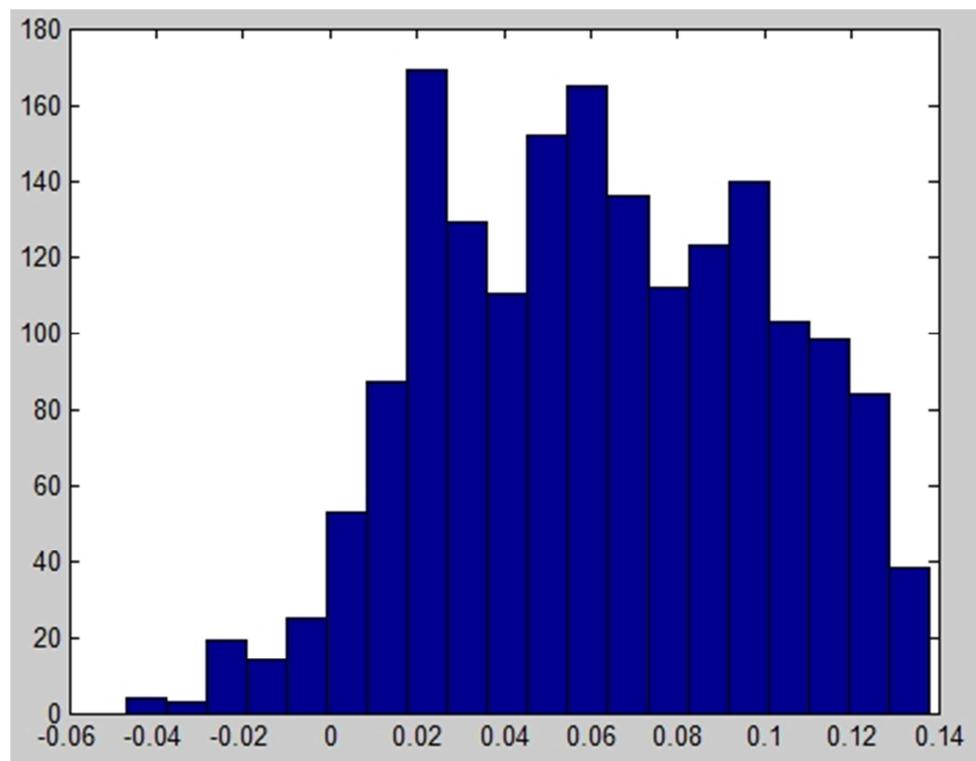


Figura 117: Histograma de los valores de la función *fitness* IS – Experimento 1, 2 y 3

periSMA	periEMA	hComienzo	hFin	Coeficiente de Determinación
200	102	0900	2100	0.3183

Tabla 81: Mejor combinación de parámetros – Experimento 1, 2 y 3

Retorno IS	Máximo drawdown IS	Función fitness IS	Retorno OOS	Máximo drawdown OOS	Función fitness OOS
3202.7	1975	0.0245	-11262	12610	-0.1812

Tabla 82: Retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 1, 2 y 3

Número total de trades	Trades en tramo alcista	Trades en tramo bajista	Trades en tramo lateral
969	568	240	161

Tabla 83: *Trades* en tramo alcista | bajista | lateral – Experimento 1, 2 y 3

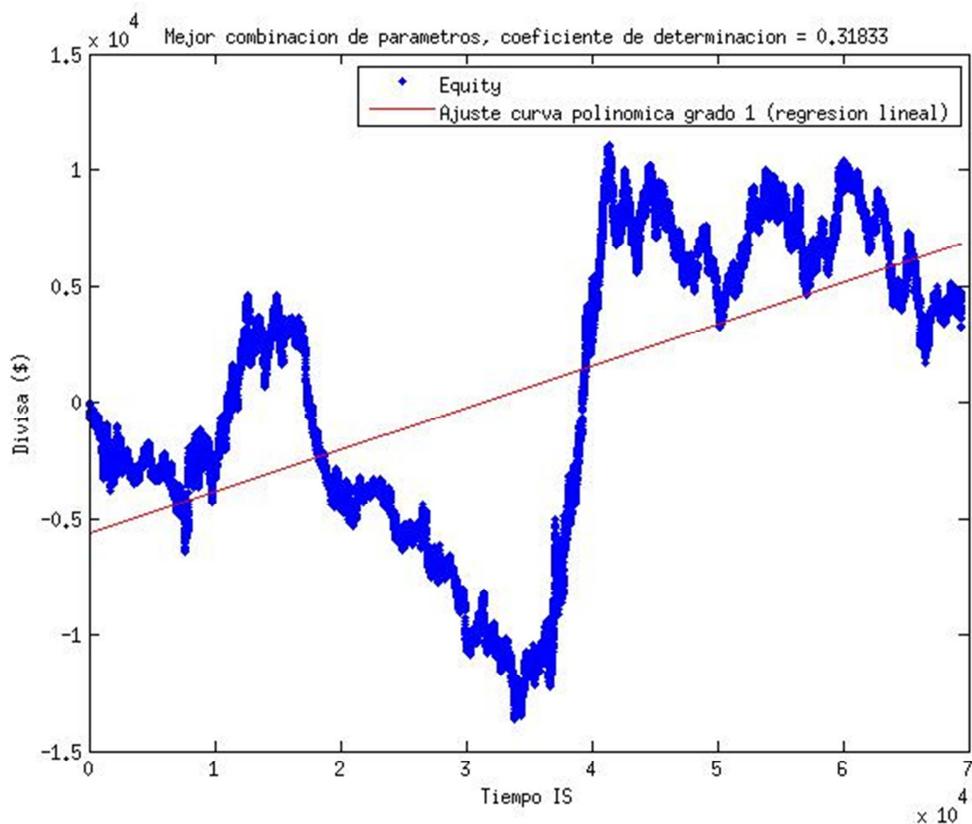


Figura 118: *Equity* IS – Experimento 1,2 y 3

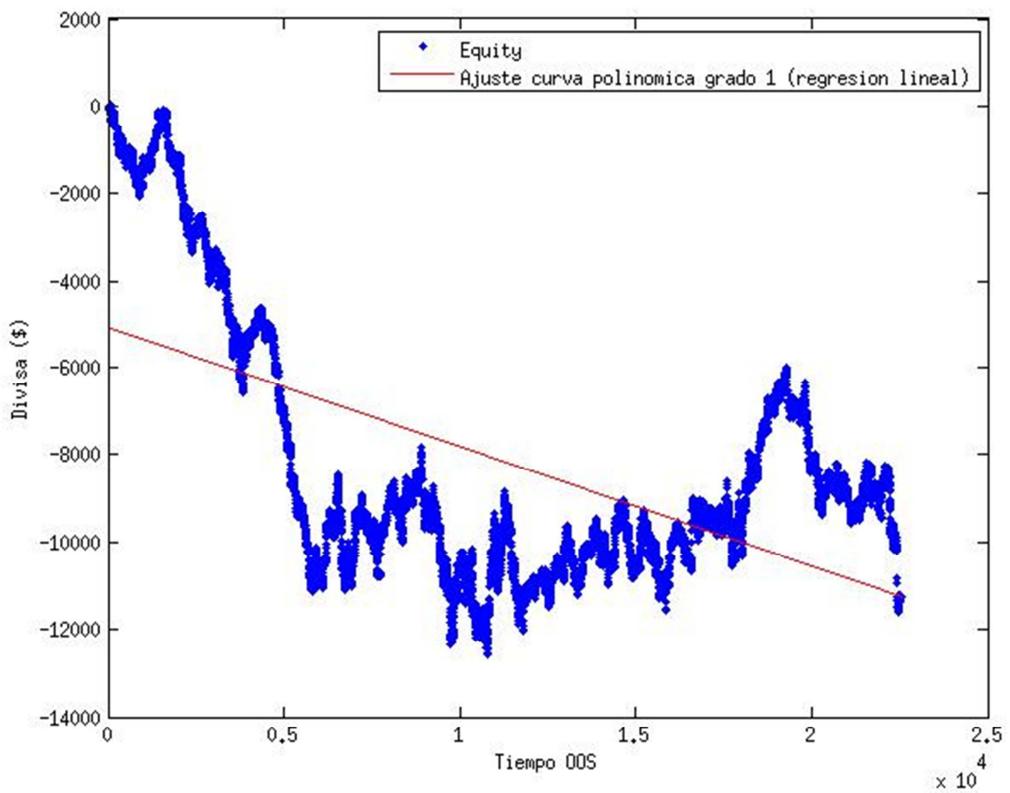


Figura 119: *Equity* OOS – Experimento 1,2 y 3

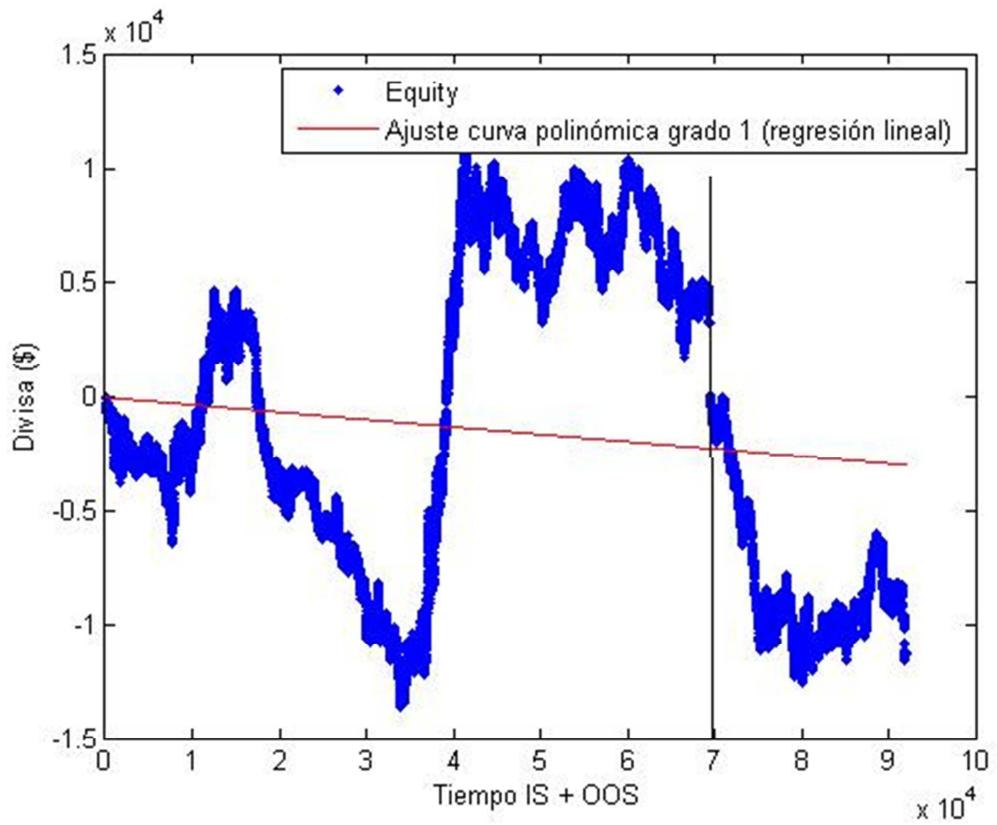


Figura 120: *Equity* IS + OOS – Experimento 1,2 y 3

Nombre del experimento: Experimento 4, 5 y 6

Número de combinaciones: 1764

Función *fitness*: SR

Filtro: 50, 150 y 300

Tiempo de ejecución de la optimización (paralelo) = 2.76 minutos

Tiempo de ejecución de la optimización (no paralelo) = 3.09 minutos

Mem [| | | | | |] 3957 / 24089 MB]

Figura 121: Consumo RAM – Experimento 4, 5 y 6

Como mínimo necesitamos una variable que ocupa $69441 * 1764 * 8 = 979.95$ MB

1 [| | 1.9%] 4 [0.0%] 7 [| | | | 100.0%] 10 [| | | | 100.0%]
2 [| | | | 100.0%] 5 [| | | | 100.0%] 8 [| 0.5%] 11 [0.0%]
3 [| | 1.4%] 6 [| 3.7%] 9 [| | | | 100.0%] 12 [| | | | 100.0%]

Figura 122: Carga del sistema – 6 cores – Experimento 4, 5 y 6

Load average: 2.51 3.58 4.01

Figura 123: Carga media del sistema – Experimento 4, 5 y 6

PBO = 100/252 (39.68 %)

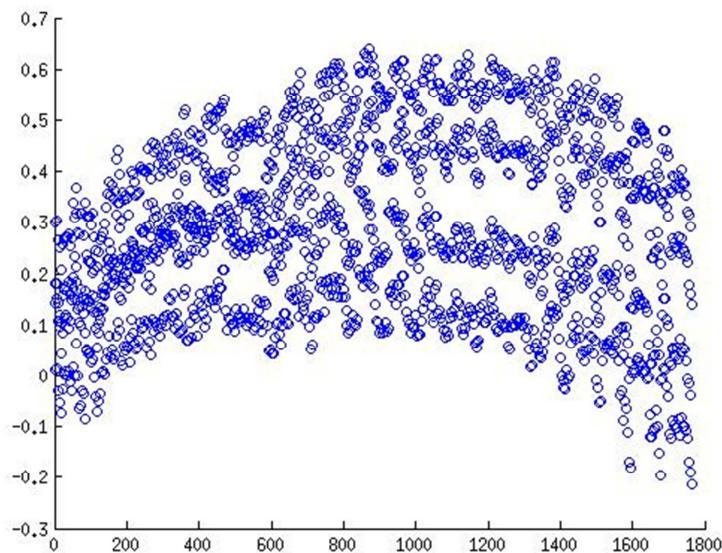


Figura 124: Nube de puntos del resultado de la función *fitness* IS – Experimento 4, 5 y 6

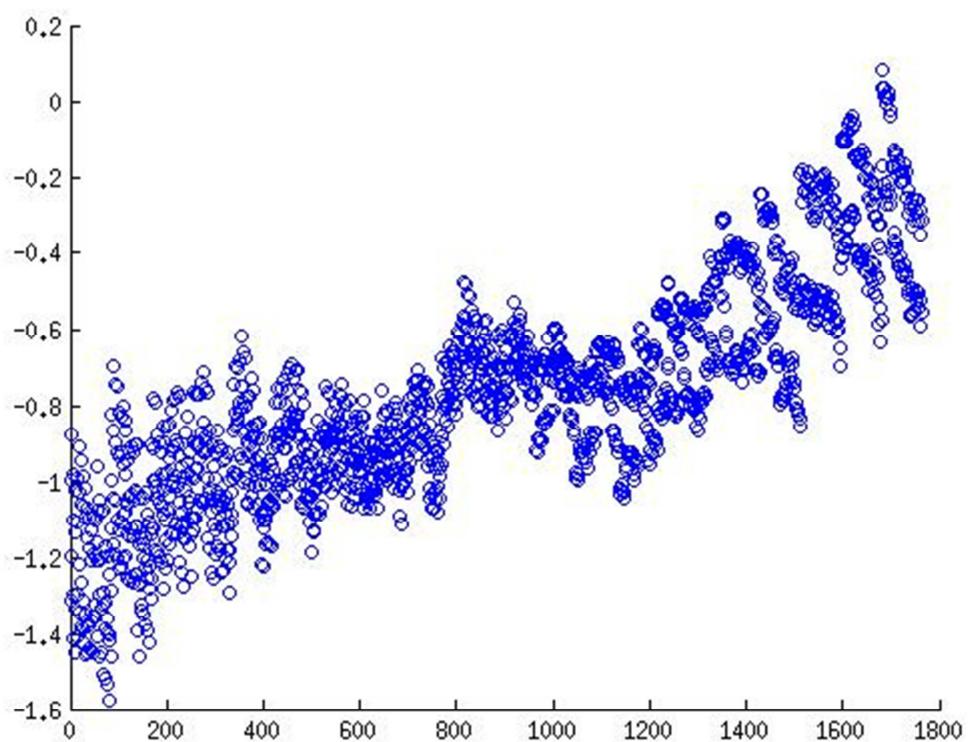


Figura 125: Nube de puntos del resultado de la función *fitness* OOS – Experimento 4, 5 y 6

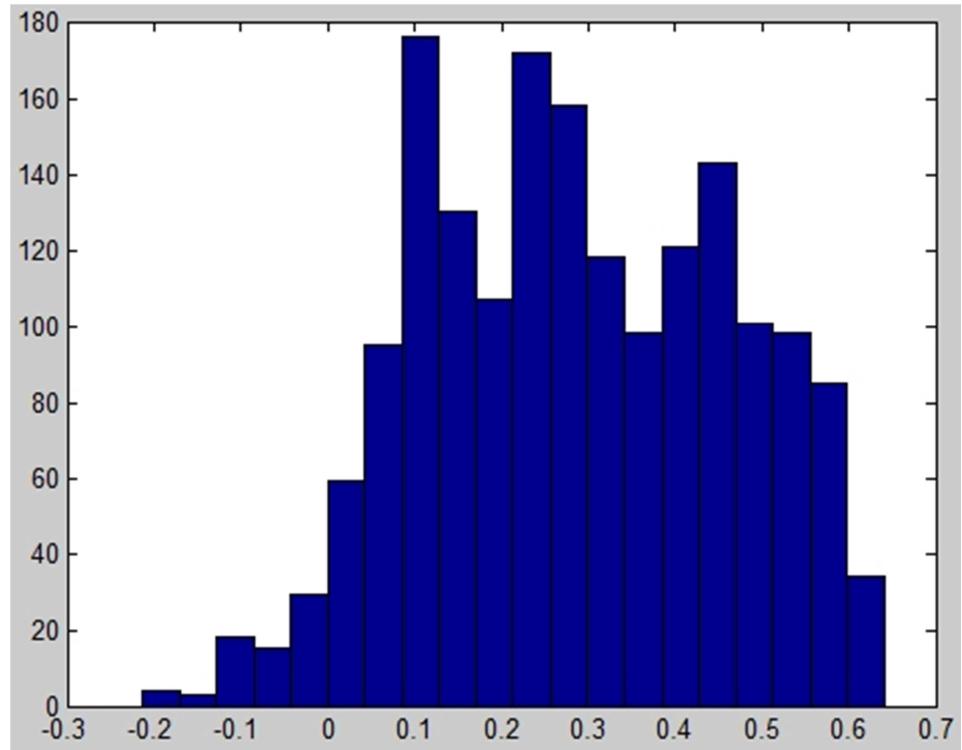


Figura 126: Histograma de los valores de la función *fitness* IS – Experimento 4, 5 y 6

periSMA	periEMA	hComienzo	hFin	Coeficiente de Determinación
200	120	0900	2100	0.3421

Tabla 84: Mejor combinación de parámetros – Experimento 4, 5 y 6

Retorno IS	Máximo drawdown IS	Función fitness IS	Retorno OOS	Máximo drawdown OOS	Función fitness OOS
3675.5	1975	0.1235	-12248	13831	-1.575

Tabla 85: Retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 4, 5 y 6

Número total de trades	Trades en tramo alcista	Trades en tramo bajista	Trades en tramo lateral
985	577	246	162

Tabla 86: *Trades* en tramo alcista | bajista | lateral – Experimento 4, 5 y 6

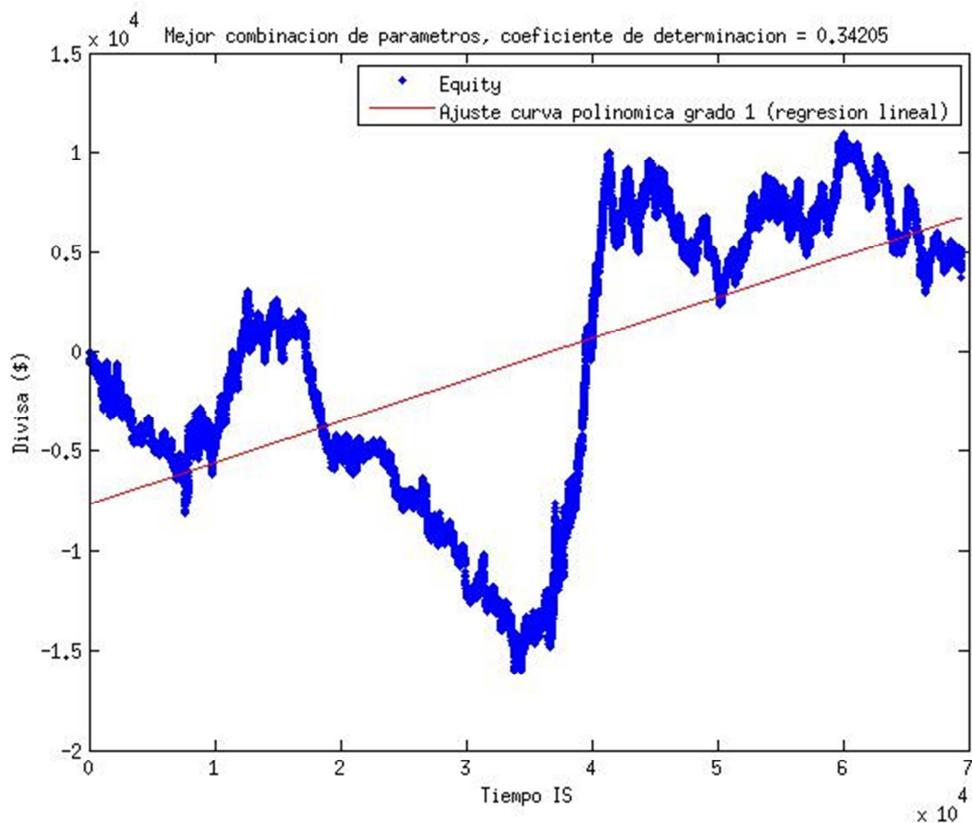


Figura 127: *Equity* IS – Experimento 4, 5 y 6

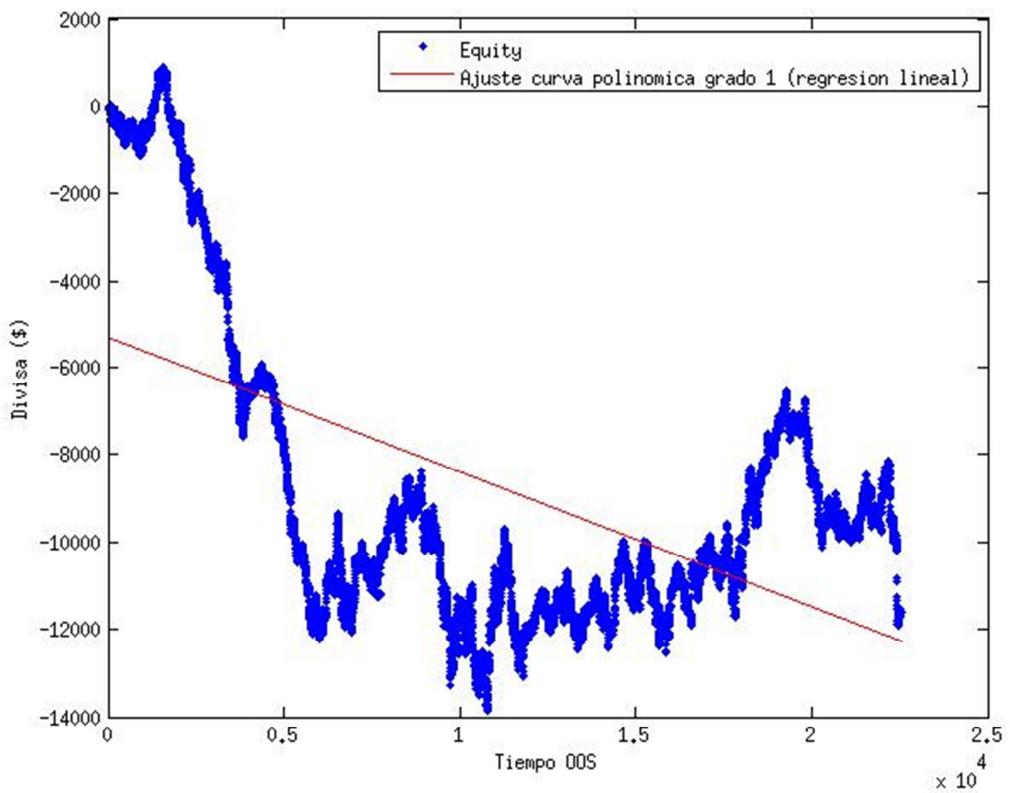


Figura 128: *Equity OOS – Experimento 4, 5 y 6*

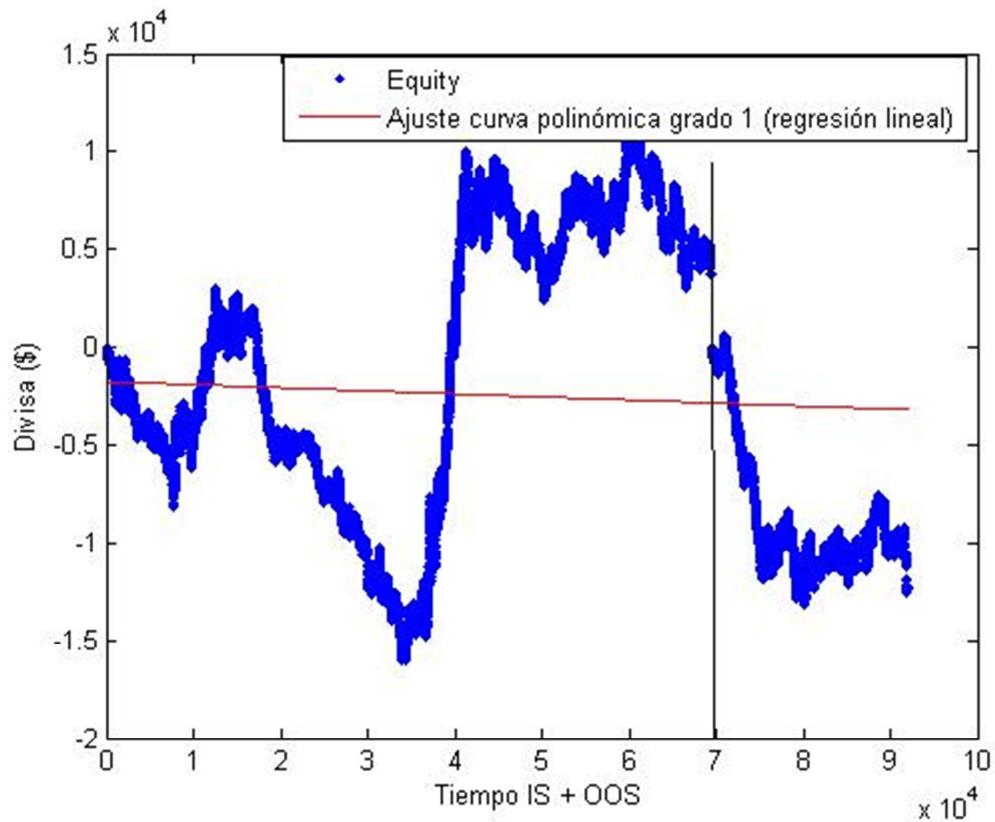


Figura 129: *Equity IS + OOS – Experimento 4, 5 y 6*

Nombre del experimento: Experimento 7, 8 y 9

Número de combinaciones: 6724

Función *fitness*: SQN

Filtro: 50, 150 y 300

Tiempo de ejecución de la optimización (paralelo) = 11.86 minutos

Tiempo de ejecución de la optimización (no paralelo) = 12.31 minutos



Figura 130: Consumo RAM – Experimento 7, 8 y 9

Como mínimo necesitamos una variable que ocupa $69441 * 6724 * 8 = 3.74$ GB

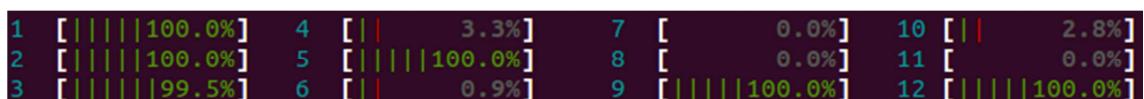


Figura 131: Carga del sistema – 6 cores – Experimento 7, 8 y 9

Load average: 6.04 5.39 4.34

Figura 132: Carga media del sistema – Experimento 7, 8 y 9

PBO = 117/252 (46.43%)

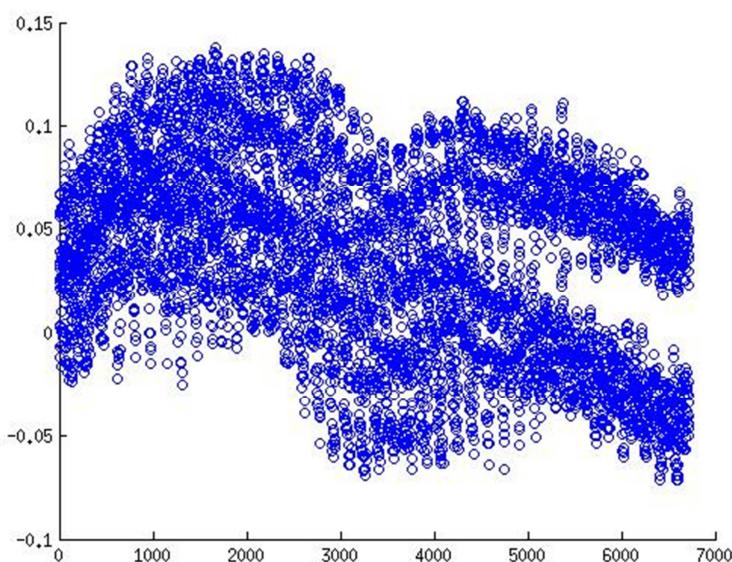


Figura 133: Nube de puntos del resultado de la función *fitness* IS – Experimento 7, 8 y 9

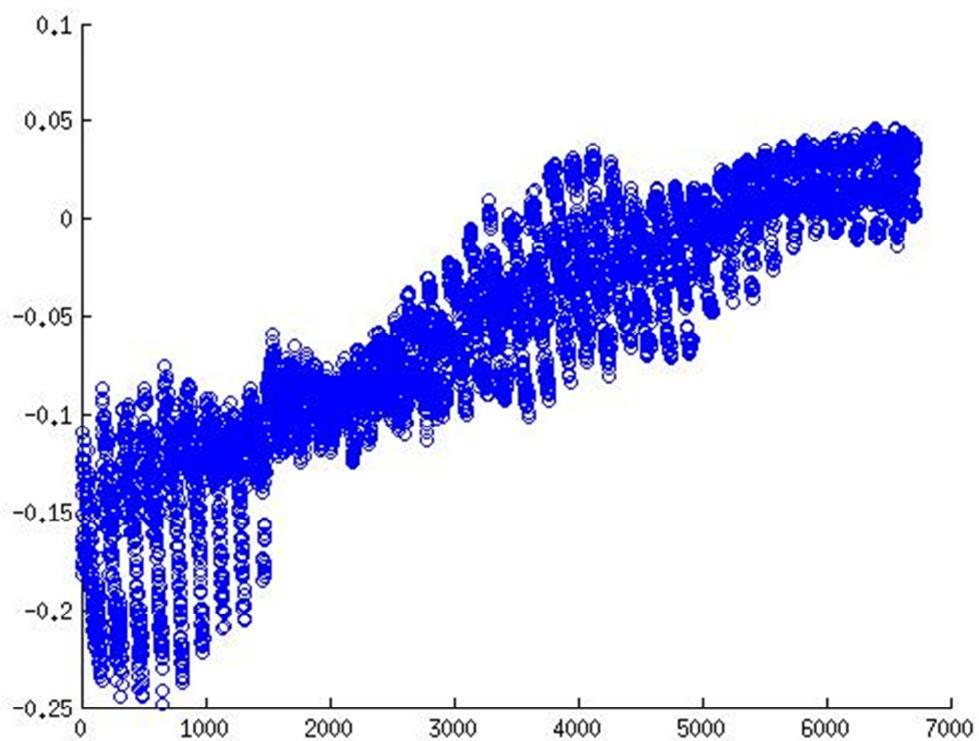


Figura 134: Nube de puntos del resultado de la función *fitness* OOS – Experimento 7, 8 y 9

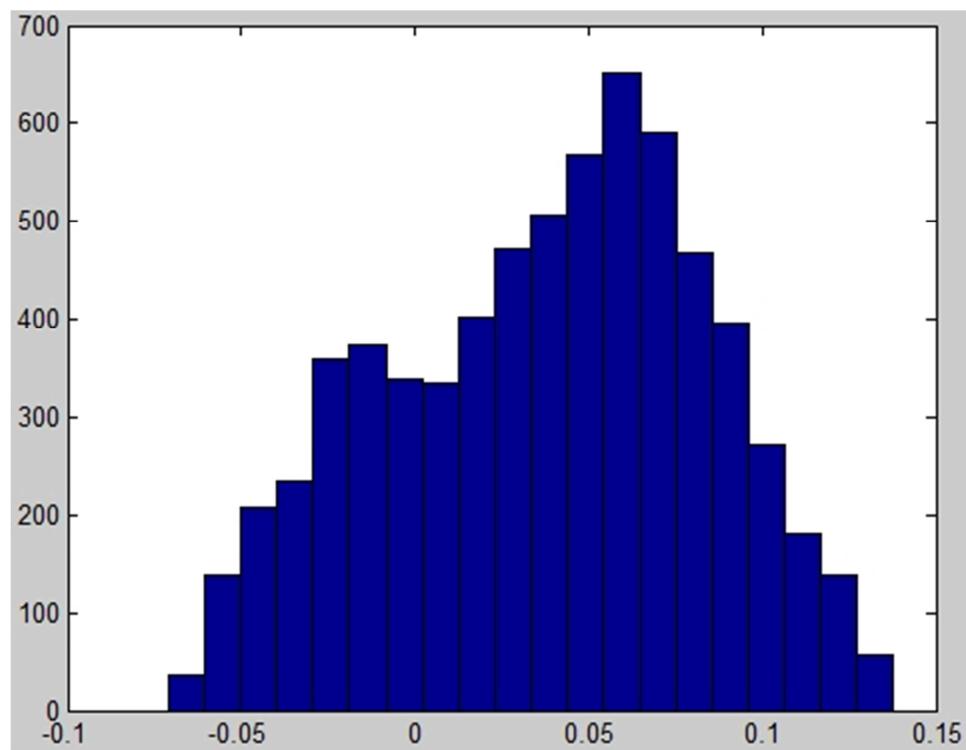


Figura 135: Histograma de los valores de la función *fitness* IS – Experimento 7, 8 y 9

periSMA	periEMA	hComienzo	hFin	Coeficiente de Determinación
200	120	0900	2200	0.5108

Tabla 87: Mejor combinación de parámetros – Experimento 7, 8 y 9

Retorno IS	Máximo drawdown IS	Función fitness IS	Retorno OOS	Máximo drawdown OOS	Función fitness OOS
8757.4	1987.5	0.0608	-11618	14754	-0.171

Tabla 88: Retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 7, 8 y 9

Número total de trades	Trades en tramo alcista	Trades en tramo bajista	Trades en tramo lateral
1003	590	246	167

Tabla 89: *Trades* en tramo alcista | bajista | lateral – Experimento 7, 8 y 9

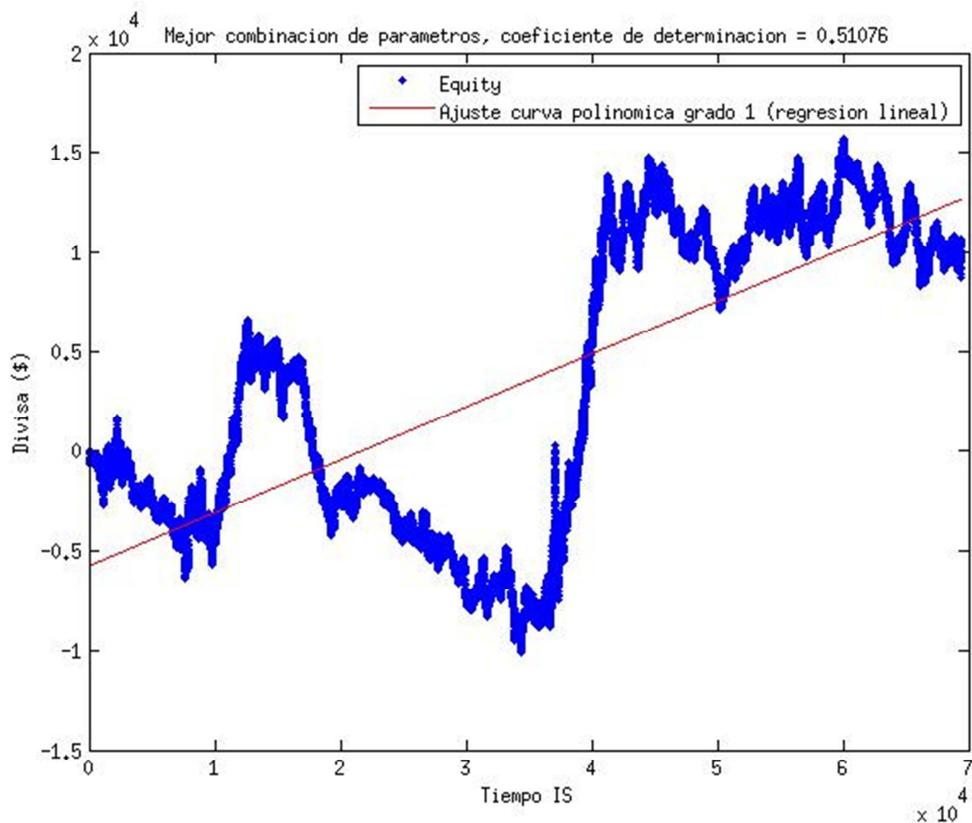


Figura 136: *Equity* IS – Experimento 7, 8 y 9

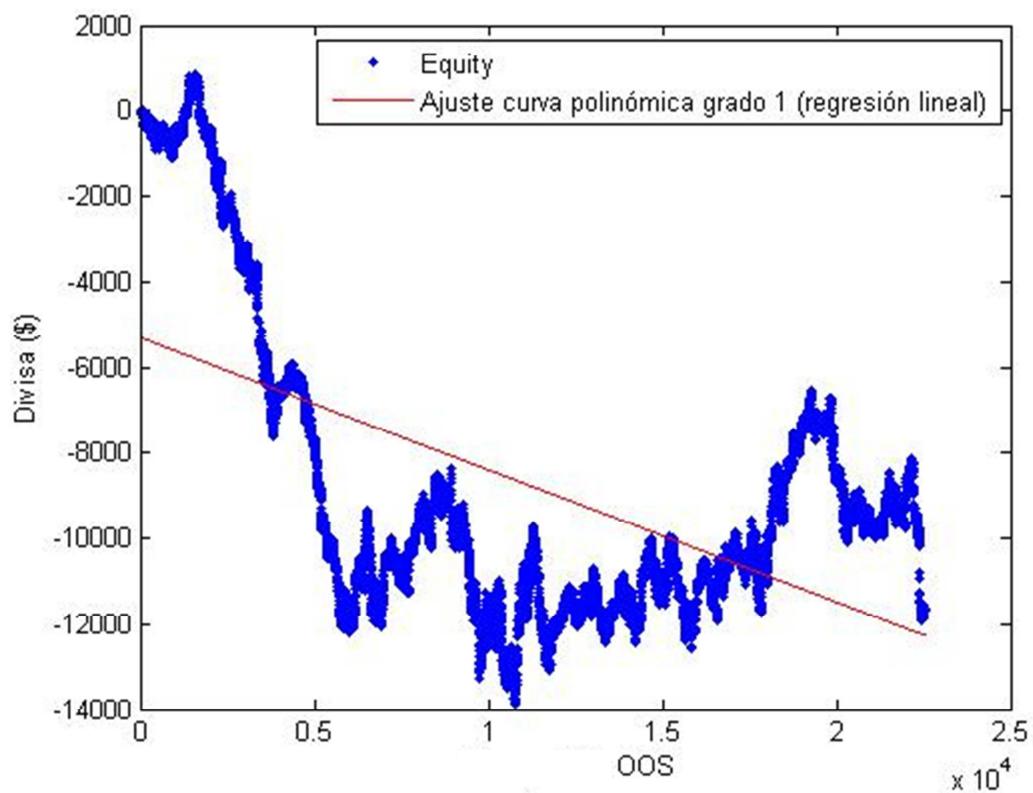


Figura 137: *Equity OOS – Experimento 7, 8 y 9*

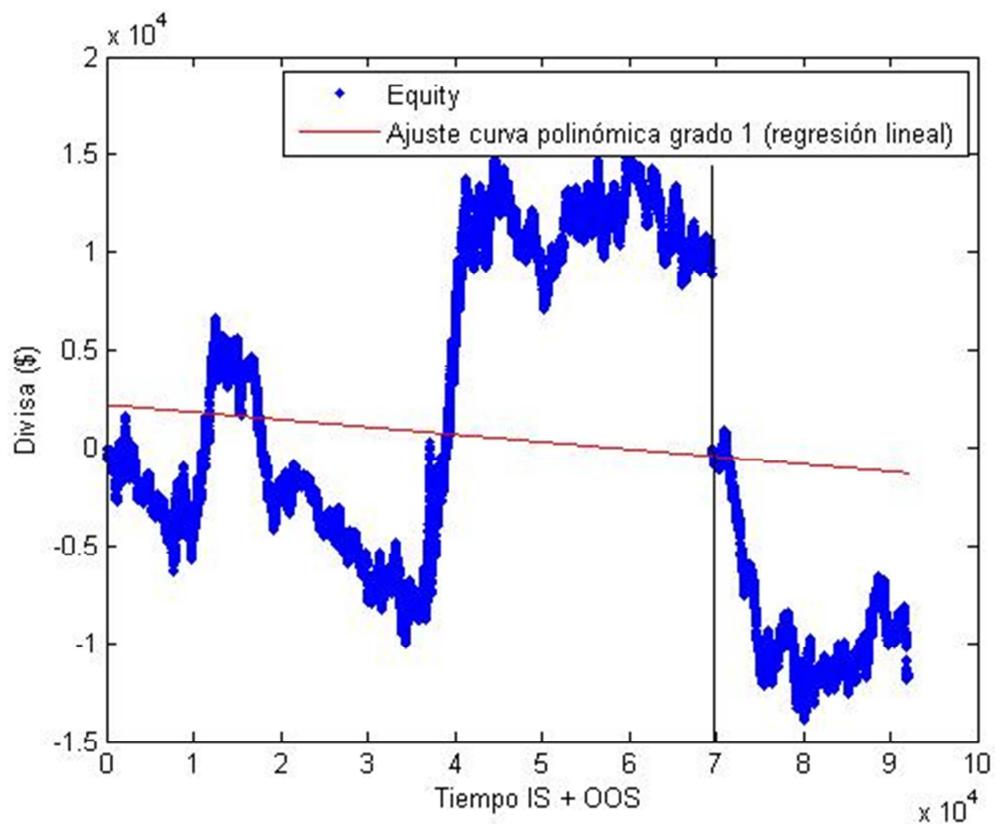


Figura 138: *Equity IS + OOS – Experimento 7, 8 y 9*

Nombre del experimento: Experimento 10, 11 y 12

Número de combinaciones: 6724

Función *fitness*: SR

Filtro: 50, 150 y 300

Tiempo de ejecución de la optimización (paralelo) = 11.86 minutos

Tiempo de ejecución de la optimización (no paralelo) = 12.31 minutos



Figura 139: Consumo RAM – Experimento 10, 11 y 12

Como mínimo necesitamos una variable que ocupa $69441 * 6724 * 8 = 3.74$ GB

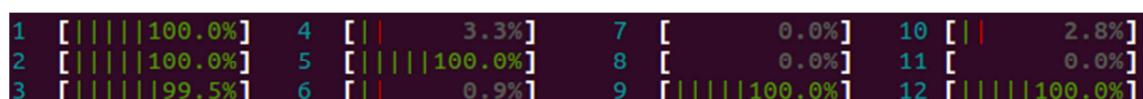


Figura 140: Carga del sistema – 6 cores – Experimento 10, 11 y 12

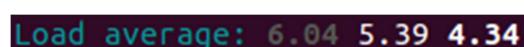


Figura 141: Carga media del sistema – Experimento 10, 11 y 12

PBO = 117/252 (46.43%)

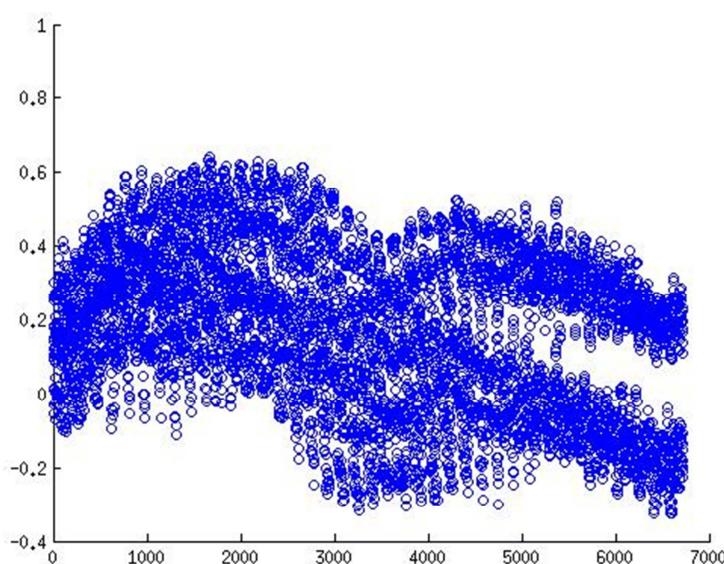


Figura 142: Nube de puntos del resultado de la función *fitness* IS – Experimento 10, 11 y 12

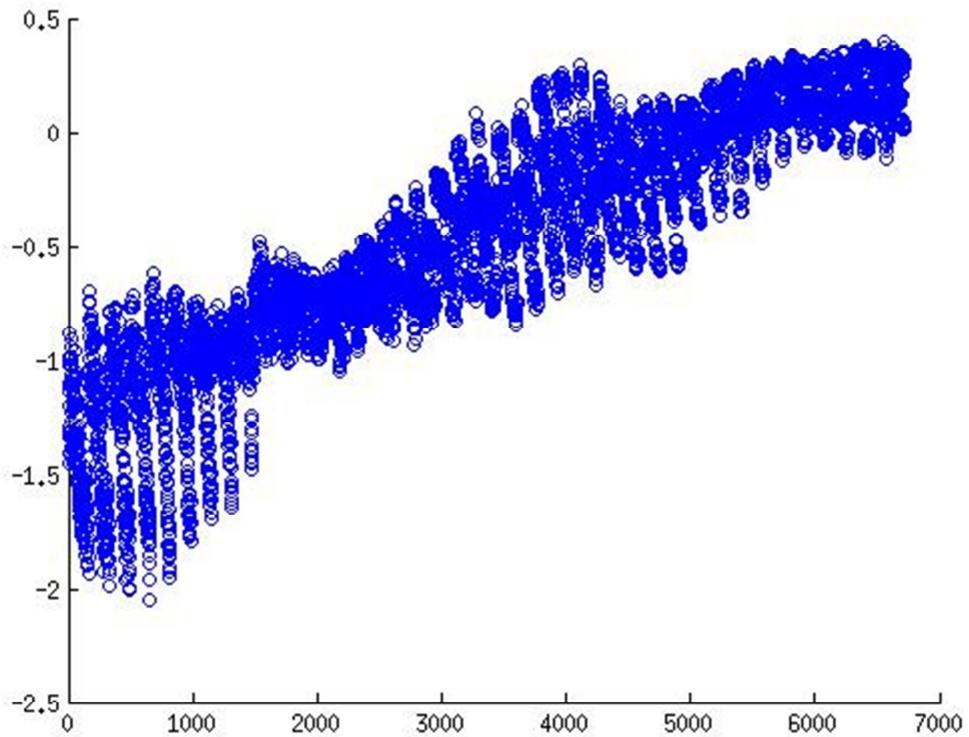


Figura 143: Nube de puntos del resultado de la función *fitness* OOS – Experimento 10, 11 y 12

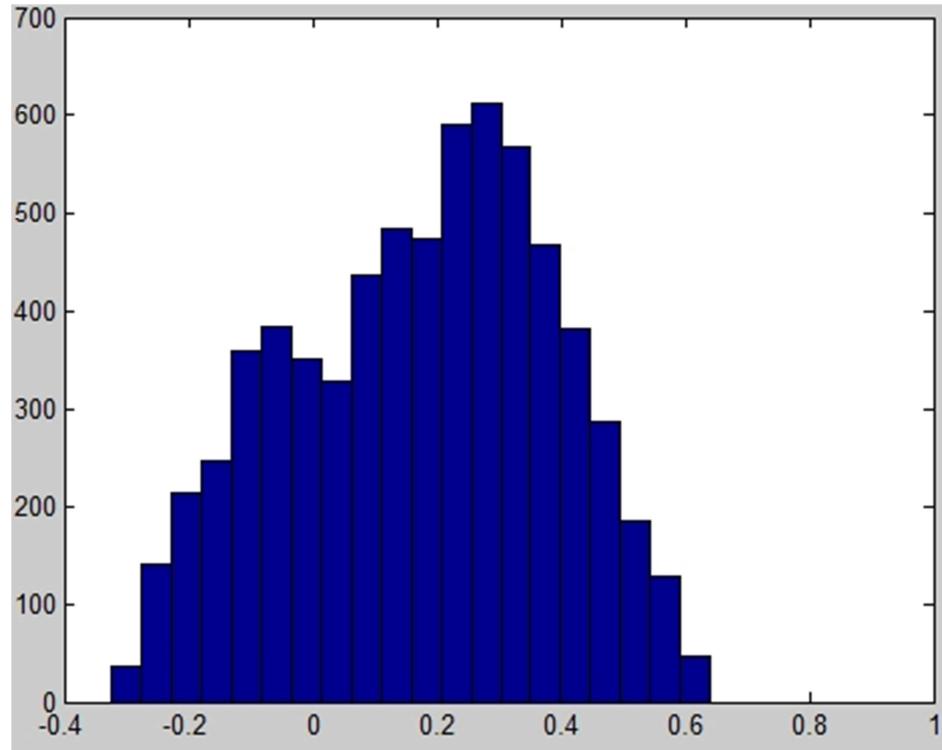


Figura 144: Histograma de los valores de la función *fitness* IS – Experimento 10, 11 y 12

Tabla 90: Mejor combinación de parámetros – Experimento 10, 11 y 12

Retorno IS	Máximo drawdown IS	Función fitness IS	Retorno OOS	Máximo drawdown OOS	Función fitness OOS
9621	1987.5	0.3010	-8275.5	12594	-0.998

Tabla 91: Retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 10, 11 y 12

Número total de trades	Trades en tramo alcista	Trades en tramo bajista	Trades en tramo lateral
995	587	242	166

Tabla 92: *Trades* en tramo alcista | bajista | lateral – Experimento 10, 11 y 12

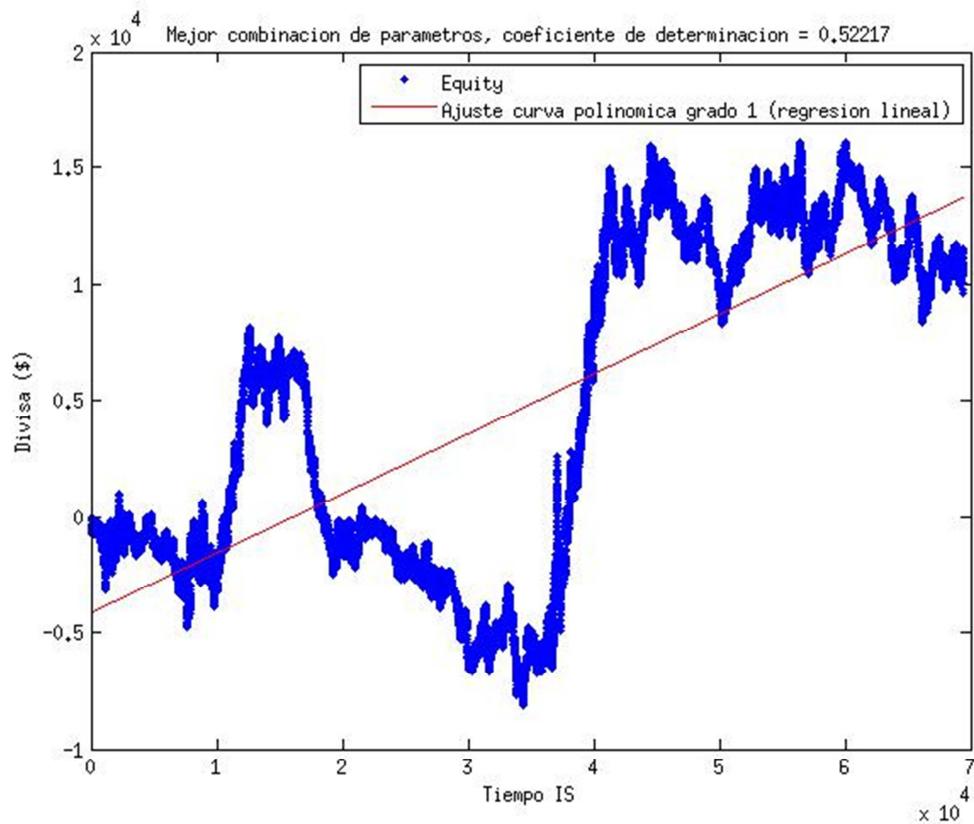


Figura 145: *Equity* IS – Experimento 10, 11 y 12

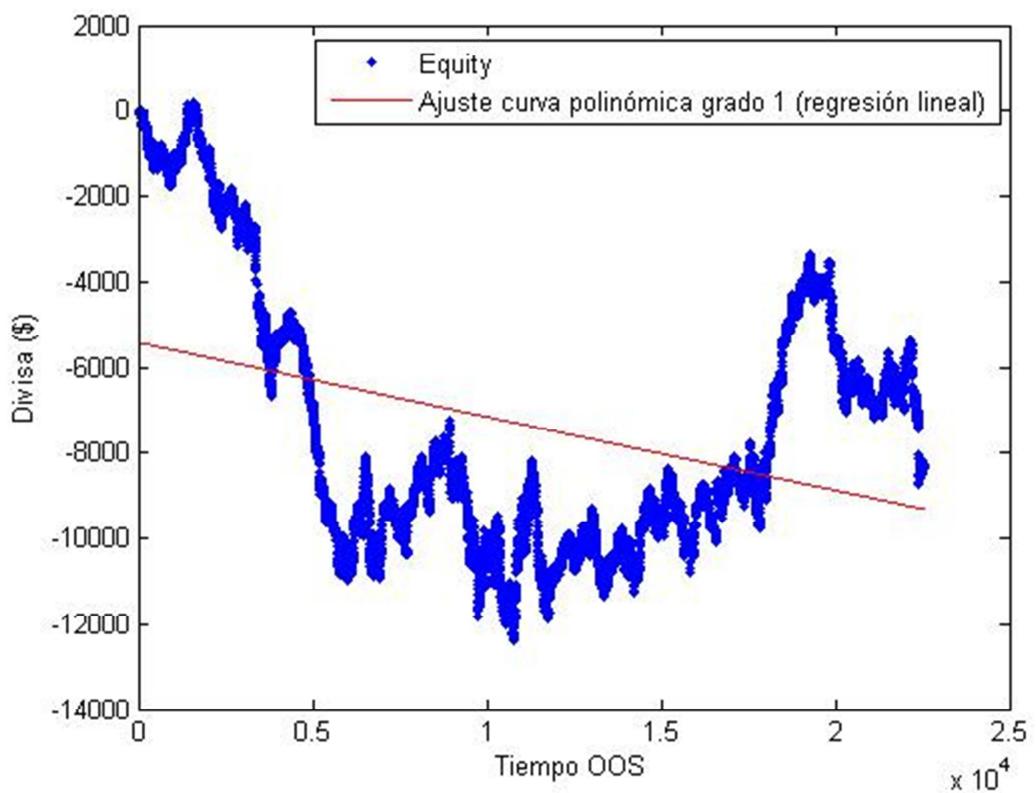


Figura 146: *Equity* OOS – Experimento 10, 11 y 12

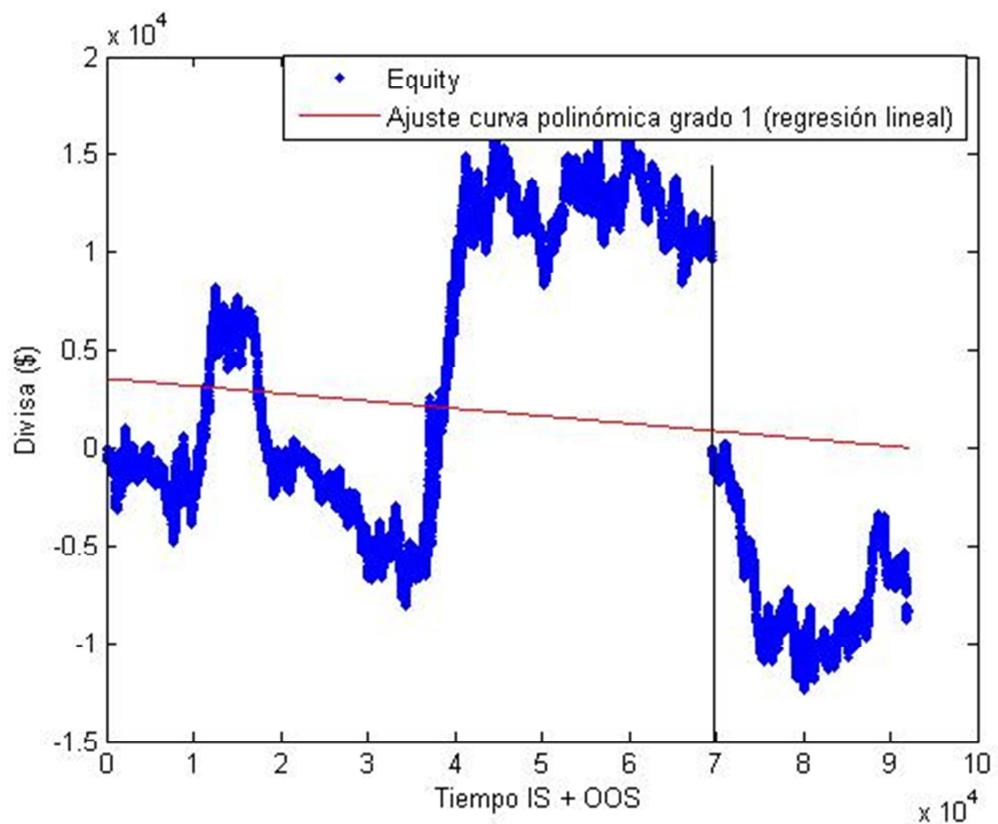


Figura 147: *Equity* IS + OOS – Experimento 10, 11 y 12

Nombre del experimento: Experimento 13, 14 y 15

Número de combinaciones: 10404

Función *fitness*: SQN

Filtro: 50, 150 y 300

Tiempo de ejecución de la optimización (paralelo) = 17.26 minutos

Tiempo de ejecución de la optimización (no paralelo) = 18.16 minutos

Mem[|||||100.0%] 8058/24089MB]

Figura 148: Consumo RAM – Experimento 13, 14 y 15

Como mínimo necesitamos una variable que ocupa $69441 * 10404 * 8 = 5.78$ GB

1 [|||||100.0%] 4 [|||||100.0%] 7 [| | 0.9%] 10 [| 0.5%]
2 [|||||100.0%] 5 [| | 1.9%] 8 [| 0.0%] 11 [| | 2.4%]
3 [|||||100.0%] 6 [| 0.0%] 9 [|||||100.0%] 12 [|||||100.0%]

Figura 149: Carga del sistema – 6 cores – Experimento 13, 14 y 15

Load average: 6.01 5.66 3.58

Figura 150: Carga media del sistema – Experimento 13, 14 y 15

PBO = 129/252 (51.2 %)

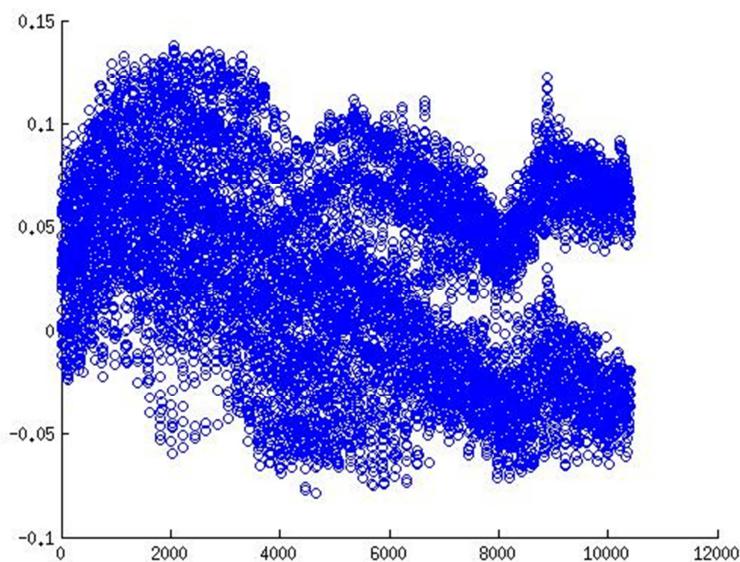


Figura 151: Nube de puntos del resultado de la función *fitness* IS – Experimento 13, 14 y 15

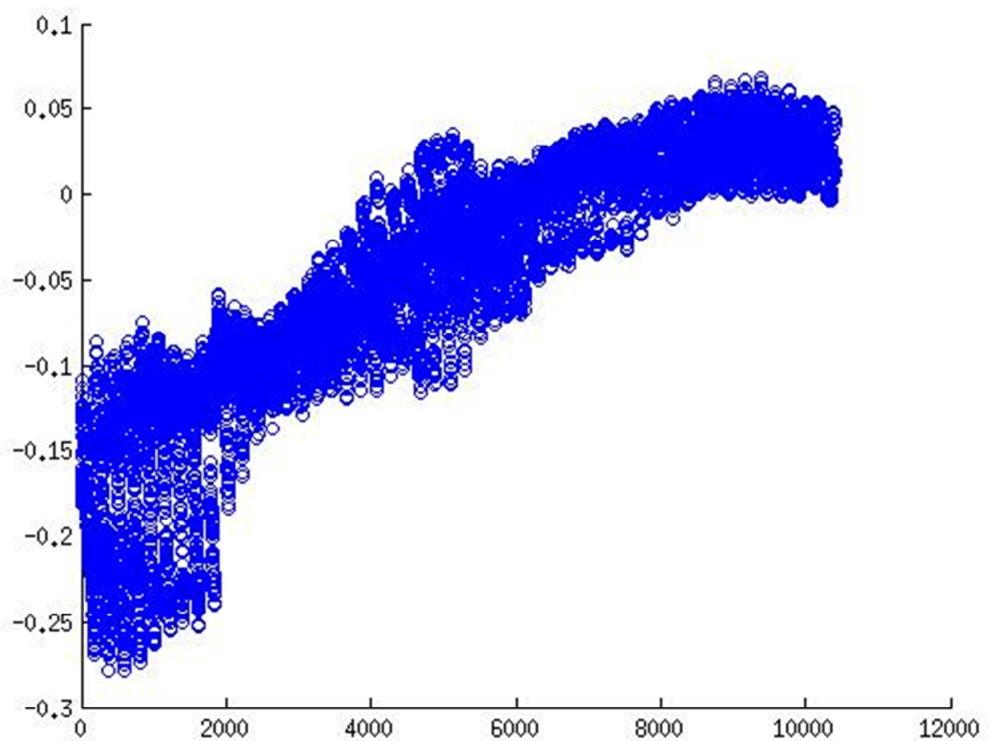


Figura 152: Nube de puntos del resultado de la función *fitness* OOS – Experimento 13, 14 y 15

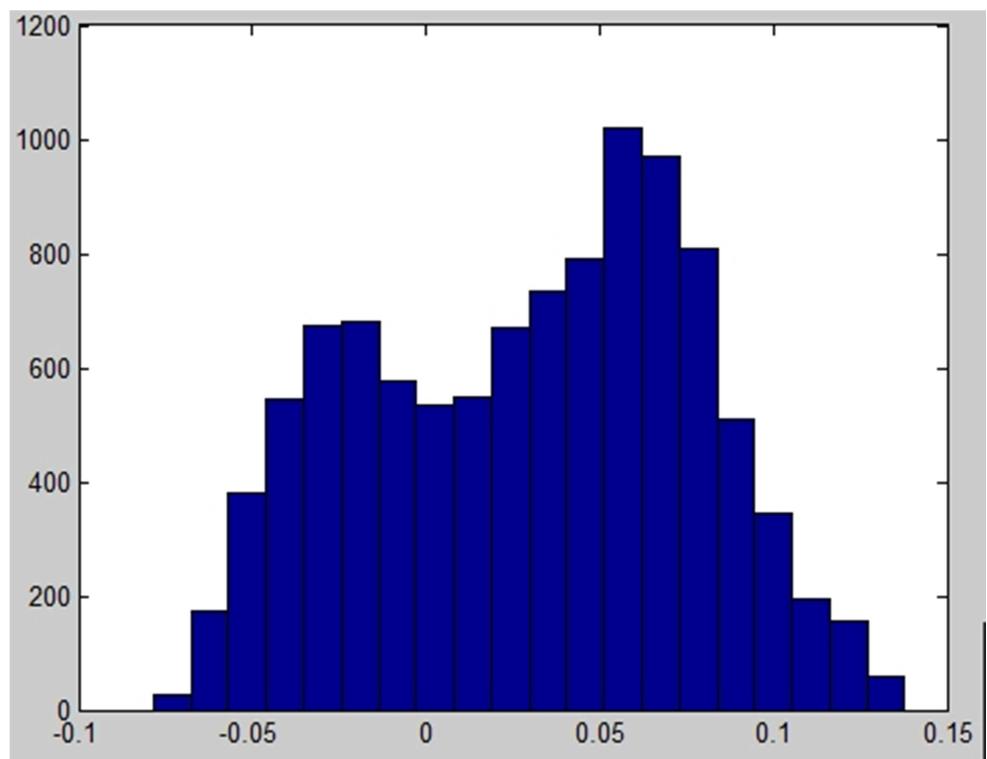


Figura 153: Histograma de los valores de la función *fitness* IS – Experimento 13, 14 y 15

periSMA	periEMA	hComienzo	hFin	Coeficiente de Determinación
200	120	0900	2200	0.5108

Tabla 93: Mejor combinación de parámetros – Experimento 13, 14 y 15

Retorno IS	Máximo drawdown IS	Función fitness IS	Retorno OOS	Máximo drawdown OOS	Función fitness OOS
8757.4	1987.5	0.0608	-11618	14754	-0.171

Tabla 94: Retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 13, 14 y 15

Número total de trades	Trades en tramo alcista	Trades en tramo bajista	Trades en tramo lateral
1003	590	246	167

Tabla 95: *Trades* en tramo alcista | bajista | lateral – Experimento 13, 14 y 15

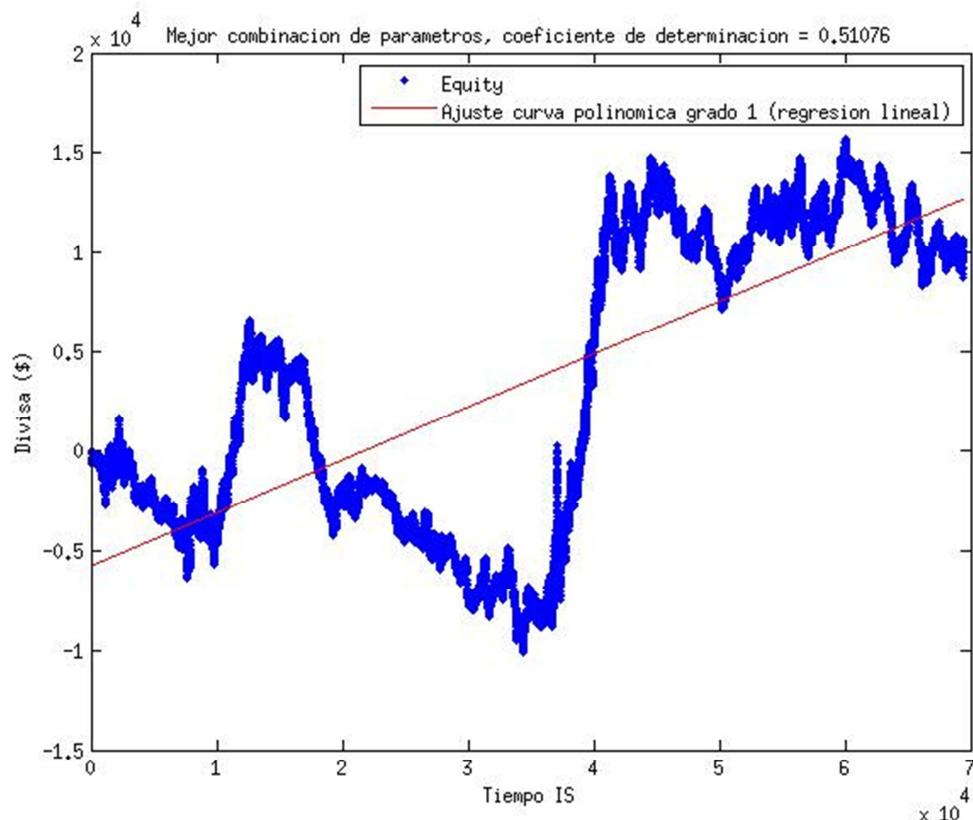


Figura 154: *Equity* IS – Experimento 13, 14 y 15

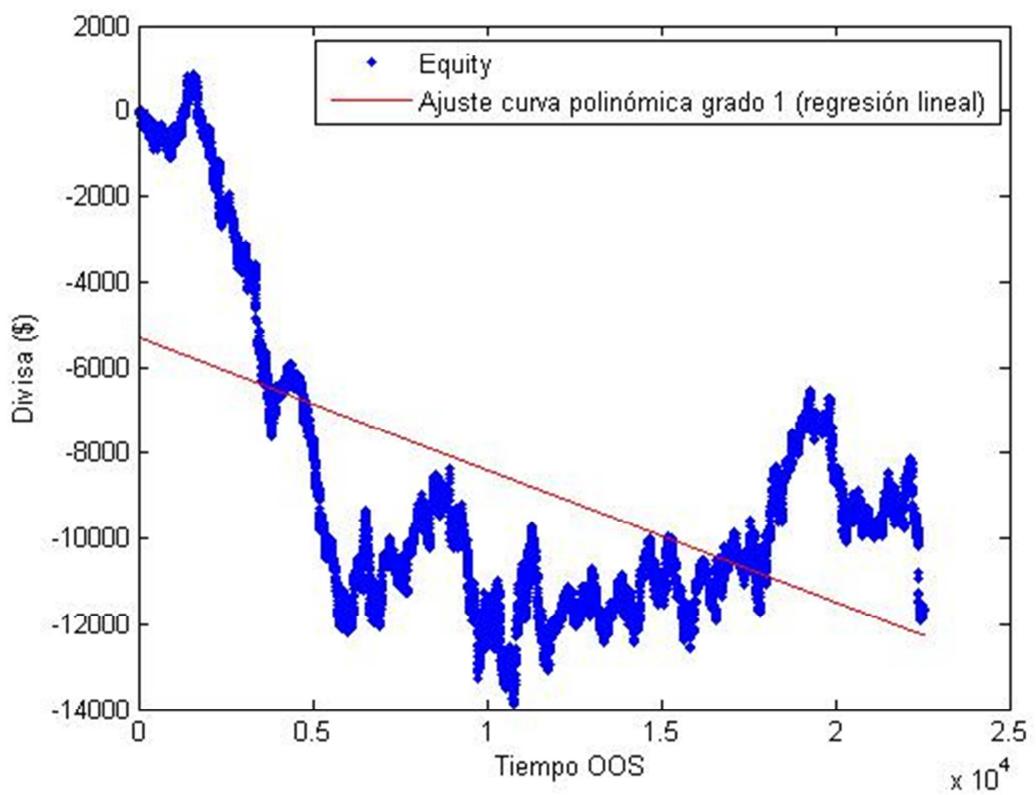


Figura 155: *Equity OOS – Experimento 13, 14 y 15*

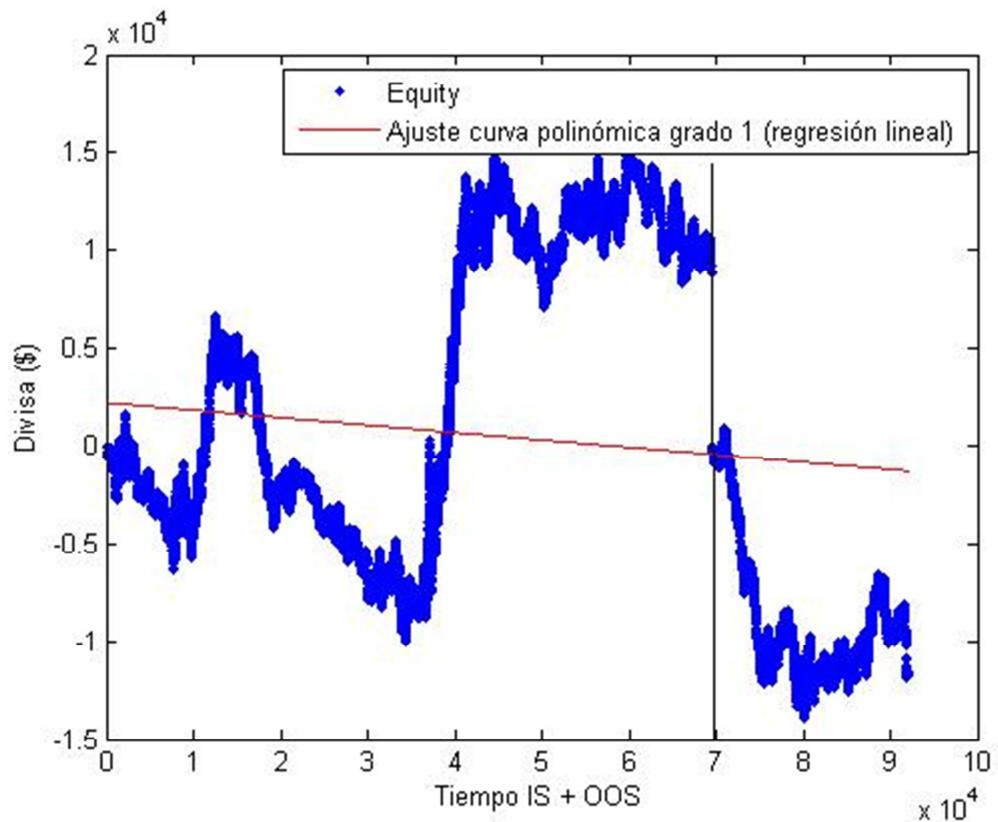


Figura 156: *Equity IS + OOS – Experimento 13, 14 y 15*

Nombre del experimento: Experimento 16, 17 y 18

Número de combinaciones: 10404

Función *fitness*: SR

Filtro: 50, 150 y 300

Tiempo de ejecución de la optimización (paralelo) = 17.26 minutos

Tiempo de ejecución de la optimización (no paralelo) = 18.16 minutos

Mem[|||||100.0%] 8058/24089MB]

Figura 157: Consumo RAM – Experimento 16, 17 y 18

Como mínimo necesitamos una variable que ocupa $69441 * 10404 * 8 = 5.78$ GB

1 [|||||100.0%] 4 [|||||100.0%] 7 [| | 0.9%] 10 [| 0.5%]
2 [|||||100.0%] 5 [| | 1.9%] 8 [| 0.0%] 11 [| | 2.4%]
3 [|||||100.0%] 6 [| 0.0%] 9 [|||||100.0%] 12 [|||||100.0%]

Figura 158: Carga del sistema – 6 cores – Experimento 16, 17 y 18

Load average: 6.01 5.66 3.58

Figura 159: Carga media del sistema – Experimento 16, 17 y 18

PBO = 129/252 (51.2 %)

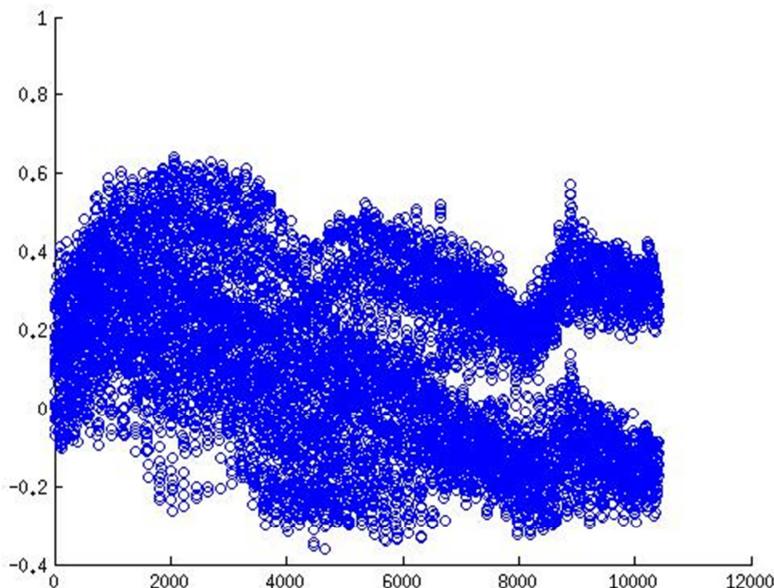


Figura 160: Nube de puntos del resultado de la función *fitness* IS – Experimento 16, 17 y 18

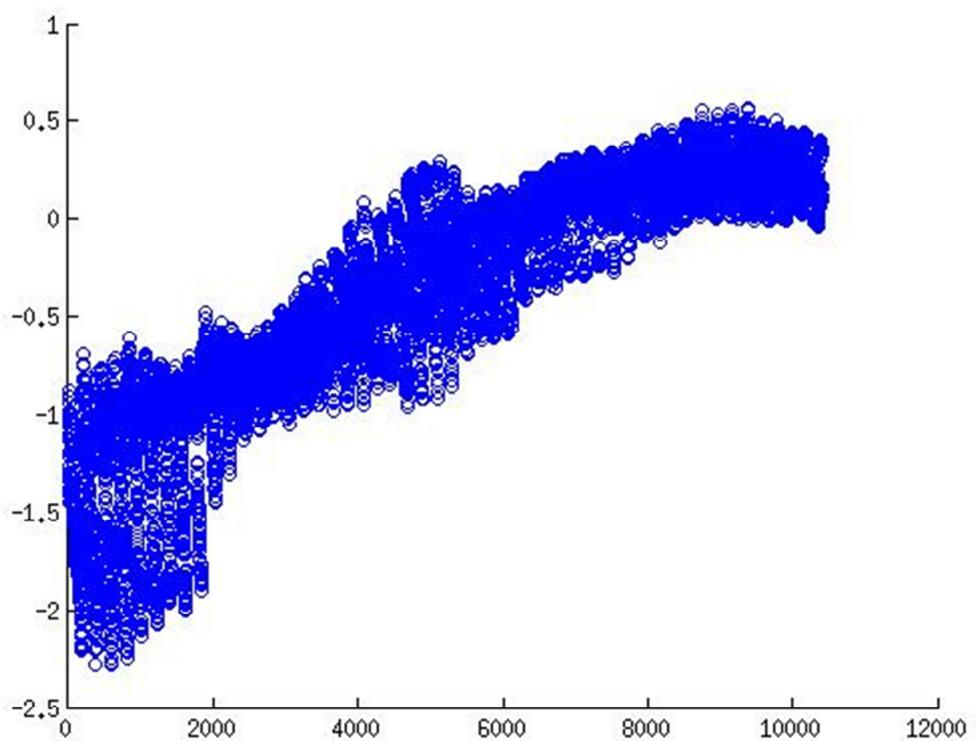


Figura 161: Nube de puntos del resultado de la función *fitness* OOS – Experimento 16, 17 y 18

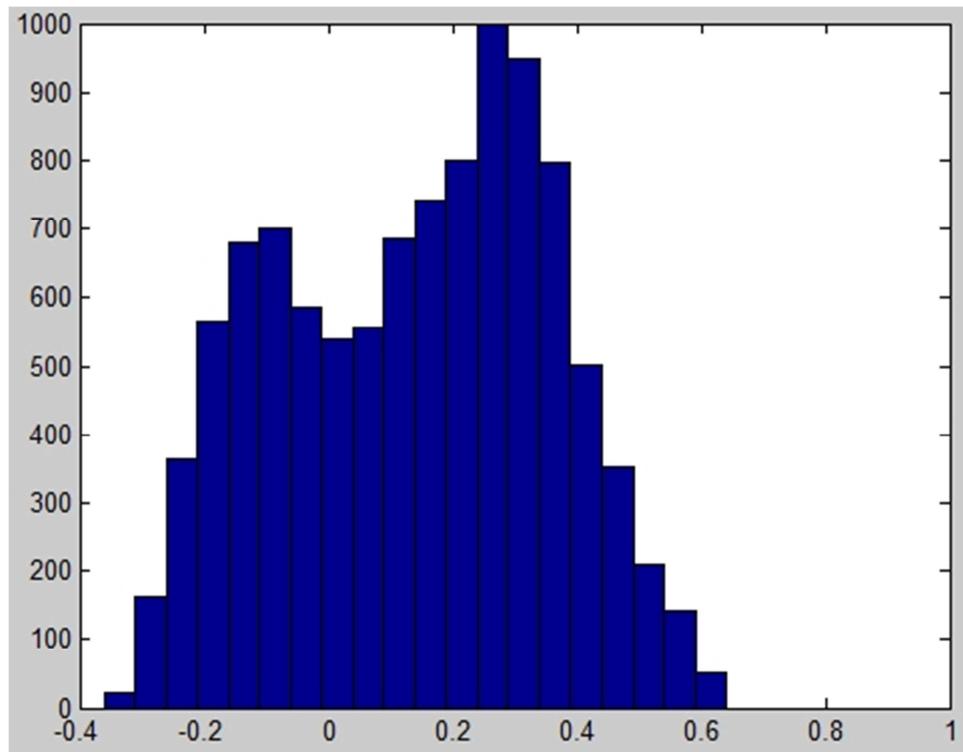


Figura 162: Histograma de los valores de la función *fitness* IS – Experimento 16, 17 y 18

periSMA	periEMA	hComienzo	hFin	Coeficiente de Determinación
200	120	0900	2200	0.5108

Tabla 96: Mejor combinación de parámetros – Experimento 16, 17 y 18

Retorno IS	Máximo drawdown IS	Función fitness IS	Retorno OOS	Máximo drawdown OOS	Función fitness OOS
8757.4	1987.5	0.2739	-11618	14754	-1.401

Tabla 97: Retorno, máximo *drawdown* y función *fitness* de la mejor combinación de parámetros en IS y OOS – Experimento 16, 17 y 18

Número total de trades	Trades en tramo alcista	Trades en tramo bajista	Trades en tramo lateral
1003	590	246	167

Tabla 98: *Trades* en tramo alcista | bajista | lateral – Experimento 16, 17 y 18

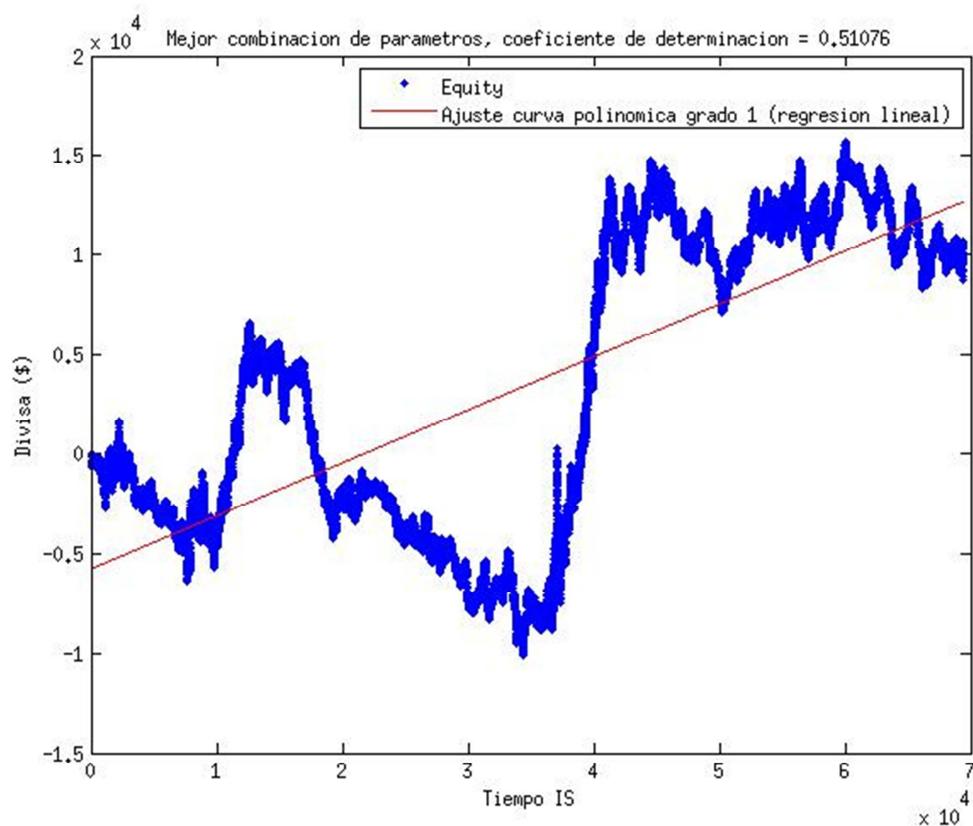


Figura 163: *Equity* IS – Experimento 16, 17 y 18

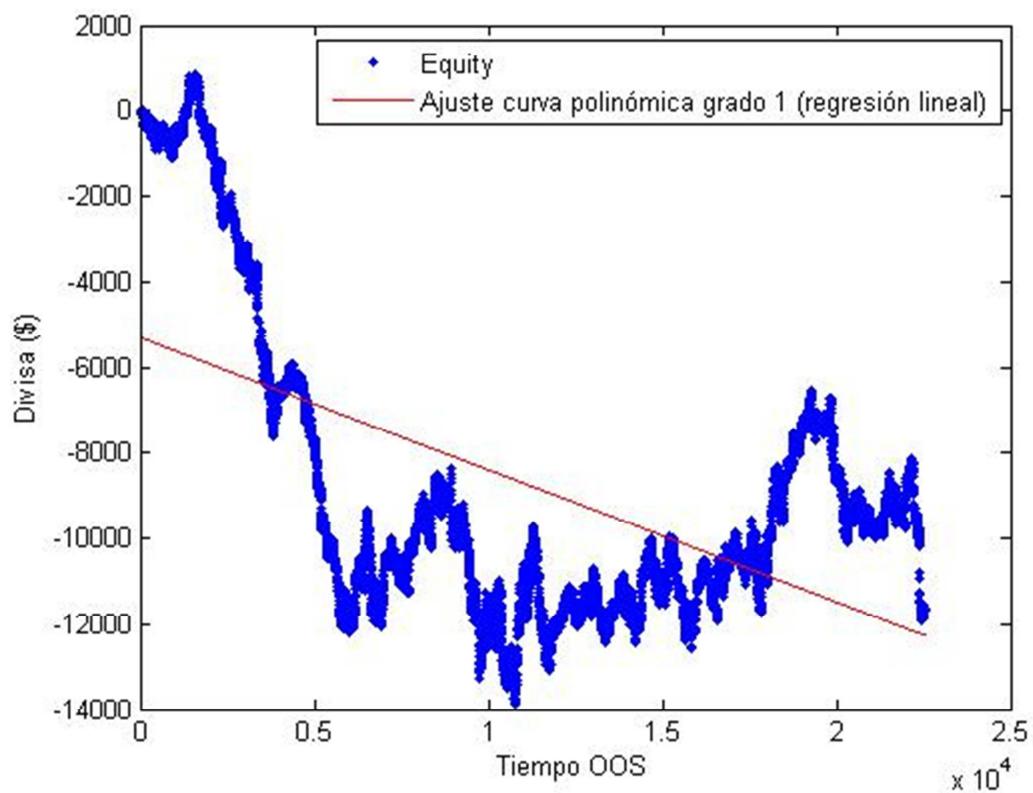


Figura 164: *Equity* OOS – Experimento 16, 17 y 18

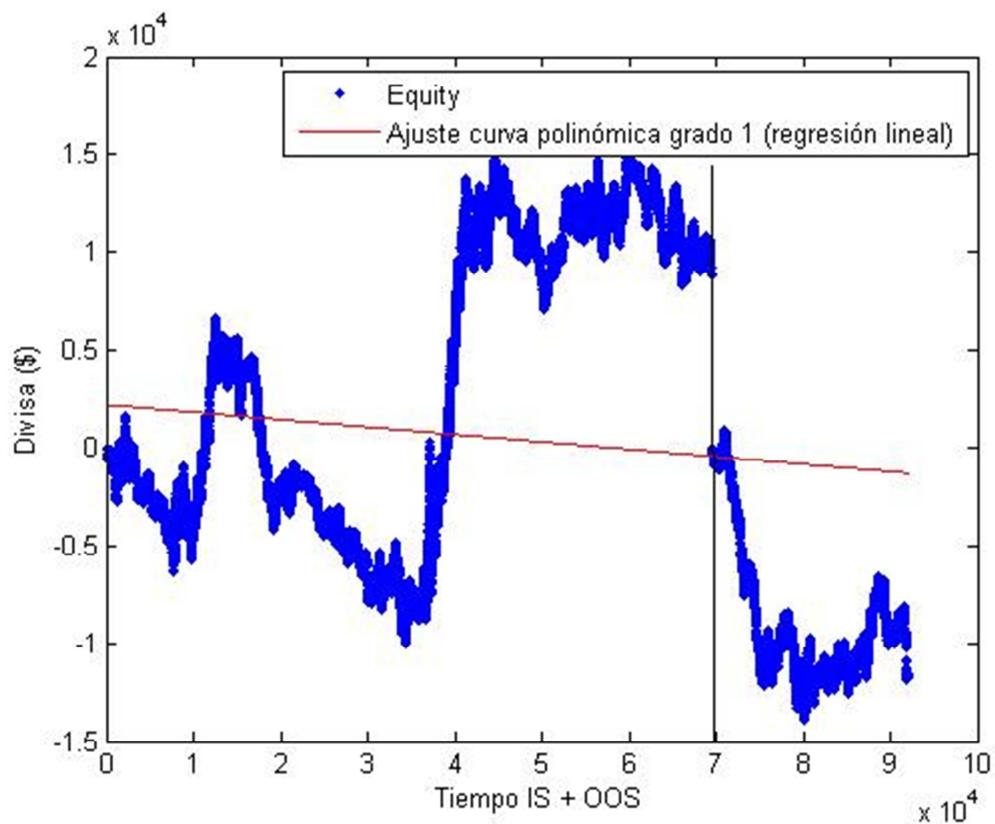


Figura 165: *Equity* IS + OOS – Experimento 16, 17 y 18

Ejemplo para comprobar la evolución de la nube de puntos OOS y confirmar la sobreoptimización.

Número de combinaciones: 40804

periSMA = 200:1:300

periEMA = 100:1:200

hComienzo = 900:100:1000

hFin = 2100:100:2200

Función *fitness*: SR

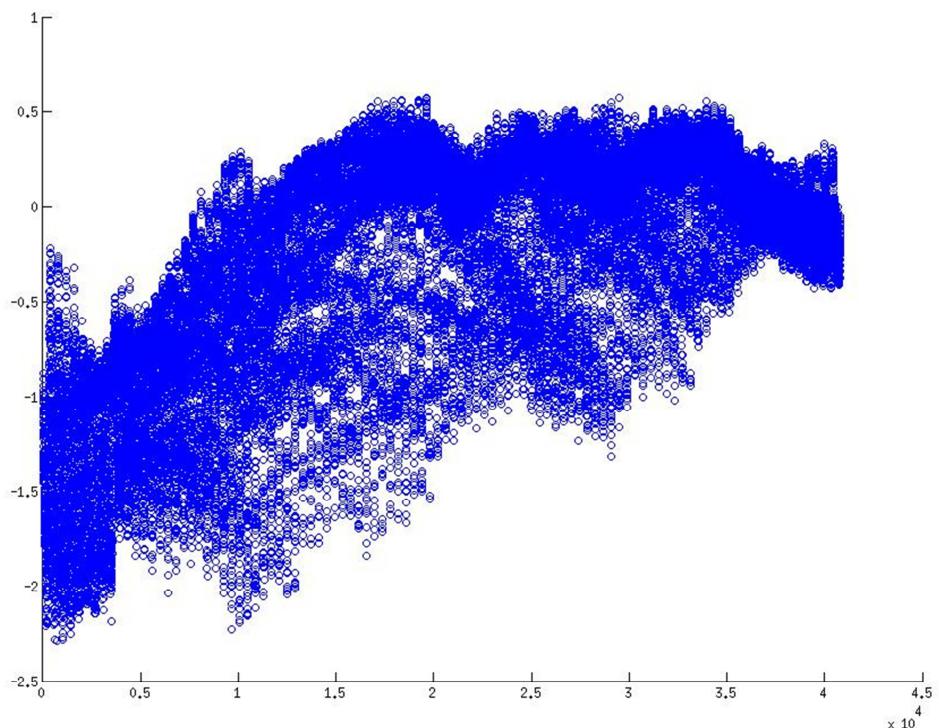


Figura 166: Nube de puntos del resultado de la función *fitness* OOS – Ejemplo de la evolución de sobreoptimización

Podemos ver que el rendimiento decrece a medida que aumentamos el número de combinaciones según los diferentes rangos en los dieciocho experimentos elaborados.

Si hubiera podido hacer el siguiente ejemplo, disponiendo de más RAM, con mayor número de combinaciones con una alta probabilidad hubiéramos visto cómo continuaría cayendo el rendimiento. Es decir, el sistema está sobreoptimizado y por ende la probabilidad de sobreoptimización aumentaría.

ANEXO IV: Informe detallado - resultados fase 3

El procedimiento de presentación de resultados será; nombre experimento, muestra de combinaciones seleccionada, función *fitness* aplicada, filtro utilizado, retorno (OOS) - máximo *drawdown* (IS y OOS) - máximo *drawdown* por el procedimiento Monte Carlo de la mejor combinación de parámetros y finalmente el Ratio Calmar.

Nombre del experimento: Experimento 1, 2 y 3

Número de combinaciones: 1764

Función *fitness*: SQN

Filtro: 50, 150 y 300

Retorno OOS	Máximo <i>drawdown</i> IS	Máximo <i>drawdown</i> OOS	Máximo <i>drawdown</i> con Monte Carlo
-11262	1975	12610	17567

Tabla 99: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 1, 2 y 3

Ratio Calmar;

$$CR = \frac{\text{Retorno OOS}}{\text{Máximo drawdown con Monte Carlo}} = \frac{-11262}{17567} = -0.64$$

Nombre del experimento: Experimento 4, 5 y 6

Número de combinaciones: 1764

Función *fitness*: SR

Filtro: 50, 150 y 300

Retorno OOS	Máximo <i>drawdown</i> IS	Máximo <i>drawdown</i> OOS	Máximo <i>drawdown</i> con Monte Carlo
-12248	1975	13831	18940

Tabla 100: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 4, 5 y 6

Ratio Calmar;

$$CR = \frac{\text{Retorno OOS}}{\text{Máximo drawdown con Monte Carlo}} = \frac{-12248}{13831} = -0.65$$

Nombre del experimento: Experimento 7, 8 y 9
 Número de combinaciones: 6724
 Función *fitness*: SQN
 Filtro: 50, 150 y 300

Retorno OOS	Máximo drawdown IS	Máximo drawdown OOS	Máximo drawdown con Monte Carlo
-11618	1987.5	14754	20743

Tabla 101: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 7, 8 y 9

Ratio Calmar;

$$CR = \frac{\text{Retorno OOS}}{\text{Máximo drawdown con Monte Carlo}} = \frac{-11618}{20743} = -0.56$$

Nombre del experimento: Experimento 10, 11 y 12
 Número de combinaciones: 6724
 Función *fitness*: SR
 Filtro: 50, 150 y 300

Retorno OOS	Máximo drawdown IS	Máximo drawdown OOS	Máximo drawdown con Monte Carlo
-8275.5	1987.5	12594	18836

Tabla 102: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 10, 11 y 12

Ratio Calmar;

$$CR = \frac{\text{Retorno OOS}}{\text{Máximo drawdown con Monte Carlo}} = \frac{-8275.5}{18836} = -0.43$$

Nombre del experimento: Experimento 13, 14 y 15
 Número de combinaciones: 10404
 Función *fitness*: SQN
 Filtro: 50, 150 y 300

Retorno OOS	Máximo drawdown IS	Máximo drawdown OOS	Máximo drawdown con Monte Carlo
-11618	1987.5	14754	20743

Tabla 103: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 13, 14 y 15

Ratio Calmar;

$$CR = \frac{\text{Retorno OOS}}{\text{Máximo drawdown con Monte Carlo}} = \frac{-11618}{20743} = -0.56$$

Nombre del experimento: Experimento 16, 17 y 18
 Número de combinaciones: 10404
 Función *fitness*: SR
 Filtro: 50, 150 y 300

Retorno OOS	Máximo drawdown IS	Máximo drawdown OOS	Máximo drawdown con Monte Carlo
-11618	1987.5	14754	20743

Tabla 104: Retorno OOS, máximo *drawdown* (IS y OOS) y máximo *drawdown* con Monte Carlo – Experimento 16, 17 y 18

Ratio Calmar;

$$CR = \frac{\text{Retorno OOS}}{\text{Máximo drawdown con Monte Carlo}} = \frac{-11618}{20743} = -0.56$$