

Título ¿Es posible superar la precisión del juicio de expertos de la estimación de esfuerzo de productos de software?

Tipo de Producto Ponencia (texto completo)

Autores Robiolo, Gabriela; Castillo, Oscar; Rossi, Bibiana y Santos, Silvana

Código del Proyecto y Título del Proyecto

A14T04 - Ingeniería de Software Empírica: prácticas que mejoran la administración de Proyectos de Software
Responsable del Proyecto

Rossi, Bibiana

Línea

Ingeniería de Software

Área Temática

TIC

Fecha

2013

¿Es posible superar la precisión basada en el juicio de expertos de la estimación de esfuerzo de productos de software?

Gabriela Robiolo

Oscar Castillo y Bibiana Rossi

Silvana Santos

Universidad Austral
Buenos Aires, Argentina
grobiolo@austral.edu.ar

Universidad Argentina de la Empresa
Buenos Aires, Argentina
oscar.alexander@gmail.com,
birossi@uade.edu.ar

Universidad Nacional de La Plata
La Plata, Argentina
silvanasantos@gmail.com

r

Abstract. La estimación de esfuerzo de productos de software basada en el juicio de expertos es el método más difundido en la industria del software y existen evidencias que puede tener la misma o mejor precisión cuando éste es comparado con modelos de estimaciones formales. Con la finalidad de brindar una mayor evidencia de esta afirmación se plantea un caso de estudio de una aplicación compleja desarrollada en el contexto de una empresa pública. Se compara el método de estimación de expertos versus los métodos formales de Regresión lineal y Analogías, utilizando modelos de una sola variable Tamaño (medido en COSMIC) o Complejidad (medida en Paths). Los resultados muestran que no fue posible superar la precisión del juicio de expertos debido a su nivel de experiencia medio-alto.

Keywords: estimación de expertos, estimación de esfuerzo, regresión lineal, analogías

1 Introducción

Jorgensen [1] afirma que la estimación de esfuerzo de proyectos de software basada en el juicio de expertos es el método más difundido en la industria del software. Si bien esta afirmación no invalida el uso de métodos formales de estimación, pone en evidencia las limitaciones de dichos métodos, los cuáles no han podido superar la capacidad humana de sintetizar diversas variables complejas. También el autor observa que la estimación de expertos puede tener la misma o mejor precisión al ser comparada con un modelo estimación formal. Encuentra una fuerte evidencia de que la

estimación de expertos es más precisa cuando el experto posee un importante conocimiento del dominio.

Se presenta en este artículo un caso de estudio con la finalidad de aportar una mayor evidencia a estas afirmaciones e investigar si es posible sostenerlas en un contexto de una empresa del ámbito público.

Las preguntas de investigación planteadas son:

- ¿Es posible reducir el error en la estimación de esfuerzo basada en expertos de un producto de software complejo aplicando un método de estimación formal que utiliza como única variable Tamaño o Complejidad?
- ¿Es posible reducir el error de estimación de esfuerzo de un producto de software complejo si a una historia sucesiva de estimaciones se aplica analogía evaluada por expertos vs analogía basada en Tamaño o Complejidad?

En segunda instancia, se optó por una sucesión de estimaciones para comprender cómo es la evaluación de analogías que realiza un experto en comparación con las analogías basadas en medidas objetivas.

Con el objetivo de responder a estas preguntas se selecciono una aplicación compleja, la cual ha sido desarrollada en sucesivas versiones, de la cual se obtuvo la especificación de requerimientos y el Esfuerzo Real (ER). También fue posible obtener la estimación de esfuerzo de un experto de la empresa para comparar sus estimaciones con los resultados obtenidos usando métodos formales de estimación.

Los métodos de estimación formales usados en la comparación son métodos frecuentemente utilizados [2]: regresión lineal [3] y analogías [4]. Se seleccionaron las métricas COSMIC [5], [6] y Paths [7], dado que la primera es un estándar internacional y la segunda una métrica de complejidad adecuada para las características de la aplicación a analizar. Además, resulta interesante que en [1] no se han incluido artículos que usen COSMIC o Paths.

COSMIC (Common Software Measurement International Consortium) function points es un estándar de medición cada vez más aceptado que mide Tamaño de requerimientos funcionales. Los requerimientos de usuarios funcionales pueden ser mapeados en procesos funcionales. Cada proceso funcional consiste en subprocesos que envuelven movimientos de datos. La cantidad de los movimientos de datos de entrada, salida, lectura y escritura es el Tamaño funcional expresado en CFP.

Paths es una medida introducida por Robiolo [7], [8] que captura la complejidad de los requerimientos. Es una aplicación de la métrica de MacCabe [9] a la descripción de requerimientos. La complejidad de los requerimientos está expresada en términos de caminos, donde cada camino corresponde a un escenario alternativo de un caso de uso y es expresado en P.

El presente artículo presenta qué se entiende por juicio de expertos y algunos de los últimos estudios empíricos en torno a este tema y una discusión bibliográfica de la conveniencia del uso de métodos de estimación basado en expertos versus métodos formales de estimación. Desarrolla un caso de estudio y finalmente detalla las conclusiones junto con la descripción de posibles trabajos futuros.

2 El juicio de expertos

Jorgensen [1], uno de los autores con mayor cantidad de publicaciones en torno a este tema en los últimos años, define una estrategia de estimación como juicio de expertos a un trabajo de estimación realizado por una persona reconocida como un experto en esta tarea, donde una parte significativa del proceso de estimación es realizada en forma intuitiva, ejecutando un proceso no explícito e inconstruible. Realizo una revisión de estudios detallada en torno a este tema. Los resultados arrojados por dicho proceso de revisión sugieren que las estimaciones basadas en juicio de expertos es la más utilizada para proyectos de software. Afirma que no existe evidencia sustancial en favor del uso de modelos de estimación y que hay situaciones donde se puede esperar que las estimaciones expertas sean mucho más precisas que los métodos formales de estimación. Además propone una lista de 12 “best practices” o principios de estimación experta validados empíricamente y provee sugerencias sobre cómo implementar estas guías en las organizaciones. Una de las mejores prácticas es buscar expertos con conocimiento del dominio y capacidad de realizar buenas estimaciones, aspecto que se destaca en este artículo.

Respecto de la evaluación de la incertidumbre de las estimaciones realizadas, se planteó el rol que cumple el feedback sobre la discrepancia existente entre las horas estimadas versus las trabajadas. Existe evidencia suficiente [10], que indica que la mayoría de los desarrolladores son, inicialmente, demasiado optimistas sobre la precisión de sus estimaciones, manteniéndose así aun cuando el feedback provisto indica lo contrario. El autor sugiere que una condición importante que se tendría que dar para mejorar las estimaciones sobre la base del feedback provisto luego de la finalización de la(s) tarea(s), sería el uso de una estrategia explícita de evaluar la incertidumbre. Esta condición mejora mientras mayor es la cantidad de información histórica de la que se dispone. Circunstancia que es confirmada en éste artículo.

Uno de los efectos negativos que más influyen en el fracaso de los proyectos de software es el exceso de optimismo. Jørgensen y Halkjelsvik [11] descubren un punto importante que pueden estar llevando a los estimadores a ser demasiado optimistas, esto es el formato utilizado al formular la pregunta que solicita la estimación de esfuerzo. El formato tradicional sería: “¿Cuántas horas se necesitan para completar la tarea X?” y el alternativo sería: “¿Cuántas tareas se pueden completar en Y horas?” Cualquiera de estos dos formatos deberían, en teoría, arrojar los mismos resultados. Según Jørgensen, cuando se utiliza el formato alternativo, se obtienen sorprendentemente estimaciones mucho más bajas y por ende mucho más optimistas que si se usara el formato tradicional. La recomendación final de dicho estudio es que siempre se opte por el formato tradicional ya que este no conlleva ninguna desviación impuesta por los clientes que quieren obtener el máximo con un presupuesto irreal. En nuestro caso de estudio se usó el formato tradicional.

Mendes [12] investigó –en el contexto de proyectos financiados por el gobierno de Nueva Zelanda y más tarde en Brasil- el uso de un enfoque centrado en el experto en combinación con una técnica que permite la inclusión explícita de la incertidumbre y las relaciones causales como medio para mejorar la estimación del esfuerzo de proyectos de software. El artículo presenta una visión general del proceso de estimación

de esfuerzo, seguido por la discusión de cómo un enfoque centrado en el experto mejora dicho procedimiento y puede ser ventajoso para las compañías de software.

3 El juicio de expertos vs los métodos formales

Jørgensen [1] presenta quince estudios que comparan la estimación de expertos con estimaciones basadas en modelos formales de estimación. Cinco de los artículos están a favor de la estimación de expertos, cinco no encuentran diferencia y cinco están a favor de los modelos formales de estimación. Resalta que el diseño de los experimentos tiene un fuerte impacto en los resultados. Además, destaca que los experimentos que no usaban modelos formales calibrados mostraban que la estimación de expertos era más precisa. Posterior al survey citado anteriormente sólo se ha hallado un artículo que compara juicio de expertos vs métodos formales, en particular analiza las ventajas y desventajas de juicio de expertos y aprendizaje automático [13].

En otro artículo, Jørgensen [14], remarca las mismas ideas resaltando que los expertos tienen una natural ventaja dado que típicamente procesan más información (o falta de ésta) en una forma más flexible y que podría ser difícil construir modelos de estimación más precisos. En el caso de estudio presentado en este artículo, se pone en evidencia la capacidad del experto anteriormente destacada.

También Jørgensen [1] reconoce que el juicio de expertos tiene sus aspectos negativos, como por ejemplo: el grado de inconsistencia y la ponderación de variables. Destaca que si estos efectos negativos se pudieran reducir, la mejora en la precisión de las estimaciones sería mucho mayor.

Jørgensen y Boehm [15], toman dos posturas opuestas y debaten intentando mostrar a las organizaciones las ventajas y desventajas de modelos formales y juicio de expertos. Boehm discierne con Jørgensen en que las estimaciones expertas producidas en estudios empíricos no son representativas de las estimaciones producidas en la práctica. Además, sostiene que ante la incertidumbre, las organizaciones van a optar por llevar a cabo extensos análisis de sensibilidad, riesgo y compensación a través de modelos formales ejecutados de manera rápida y con muy poco esfuerzo humano. Boehm recomienda combinar ambos enfoques, almacenando los resultados al finalizar el desarrollo (o fases del mismo) y utilizar estos valores en el proceso de “close-loop feedback” donde se comparan las entradas de las estimaciones, las salidas y los resultados finales de manera de aprender de eso y poder calibrar estimaciones de futuros proyectos. Jørgensen alienta a las organizaciones a invertir en estudios e investigaciones para mejorar los procesos de estimación basados en juicio experto. Esto mismo afirman [12][16]. También Jørgensen propone el uso del método Wideband Delphi de Bohem para mejorar las estimaciones y evitar posibles pujas de intereses entre los stakeholders.

MacDonell and Shepperd [17], afirman que hay un alto grado de interdependencia entre las estimaciones basadas en los modelos comunes de estimación y estimación de expertos, y es difícil derivar reglas para seleccionar el modelo de estimación más preciso, por lo tanto la solución parece ser usar la combinación de modelos.

4 Caso de Estudio

La empresa donde se desarrolló el caso de estudio es una empresa grande y compleja del ámbito público argentino. La aplicación desarrollada es un sistema de infracciones de tránsito. Por motivos de confidencialidad no se describen más aspectos. La especificación de requerimientos de la aplicación seleccionada se basó en casos de uso (CU). Fueron seleccionados los casos de uso de cinco versiones que estaban claramente documentados, la codificación había sido finalizada y tenían registrado el ER. Se realizó la medición de Tamaño funcional y Complejidad de los casos de uso aplicando las métricas COSMIC y Paths respectivamente.

La Tabla 1 muestra para cada caso de uso el ER medido en horas hombre, el Tamaño medido en COSMIC, la Complejidad medido en Paths y el valor de las horas estimadas por un experto expresado en horas hombre para los CU de la aplicación y para las estimaciones sucesivas de las versiones.

Table 1. Datos de la aplicación del ámbito público

Versión	ID CU	ER	CFP	P	Experto (CU)	Experto (versiones sucesivas)
7	1	264	20	3	240	240
7	2	32	5	3	24	24
7	3	248	15	23	208	208
7	4	112	37	15	88	88
9	5	104	16	5	80	80
9	6	136	15	9	96	96
10	7	56	10	3	64	64
10	8	184	7	25	112	120
10	9	416	93	63	328	344
10	10	8	11	2	8	16
10	11	16	4	2	8	16
11	12	208	90	58	176	200
11	13	24	5	2	16	16
11	14	144	149	16	120	144
11	15	96	54	5	96	88
12	16	520	46	43	504	504
12	17	112	97	6	136	144
12	18	40	71	30	40	40

En la Tabla 2 y 3 se muestran las características del experto que realizó las estimaciones, las cuales describen el perfil, el nivel de experiencia y el grado de conocimiento del entorno del experto. Se le pidió al experto que clasificara sus capacidades en uno de los siguientes niveles: alto, medio o bajo. El experto es una persona que tiene un nivel alto de experiencia, tanto en el desarrollo de software como en el liderazgo, estimación de proyectos y en la tecnología usada en el proyecto. En el momen-

to que realizó las estimaciones el experto estaba a cargo del sector donde se desarrolló la aplicación, pero no trabajaba en este sector cuando se desarrollaron las versiones que se muestran en este caso de estudio, por lo tanto no tenía conocimiento del ER. A solicitud de los autores del artículo realizó la estimación, basándose en la especificación de requerimientos, sin conocer las horas reales de la aplicación.

Table 2. Perfil del experto

Aptitudes	Descripción
Título de grado	Ingeniero
Años de experiencia en el desarrollo de software	12
Años de experiencia en liderazgo	8
Especialidad, Conocimientos	Desarrollo .NET, Java patrones, liderazgo de equipos, metodologías de desarrollo

Table 3. Nivel de experiencia y grado de conocimiento del entorno del experto

Capacidades	Alto	Medio	Bajo
Nivel Experiencia	x		
Conocimiento del rendimiento de los perfiles del grupo de desarrollo		x	
Conocimiento de la Tecnología	x		
Conocimiento del Dominio		x	

Para el cálculo de los errores se utilizara la Magnitud Relativa del Error (MRE) y el error relativo (ErR) siguiendo las fórmulas 1 y 2, respectivamente.

$$MRE = \text{abs}(ER - EE) / ER \quad (1)$$

$$ErR = (ER - EE) / ER \quad (2)$$

Además, la calidad de la predicción (PQ) se calculará aplicando la siguiente fórmula (3)

$$PQ(0.25) = k/n \quad (3)$$

Siendo k la cantidad de CU cuyo error es menor a 0.25 y n el total de los casos de uso [18].

Con el objetivo de testear estadísticamente los resultados se plantean las siguientes hipótesis alternativas:

- H1_a: La media del MRE de la estimación de experto es menor que la media del MRE de la estimación usando el modelo de regresión lineal y la variable independiente P.

- H1_b: La media del MRE de la estimación de experto es menor que la media del MRE de la estimación por Analogía medido el Tamaño en CFP.
- H1_c: La media del MRE de la estimación de experto es menor que la media del MRE de la estimación por Analogía medida la Complejidad en P.
- H1_d: La media del MRE de la estimación de experto es menor que la media del MRE de la estimación por Analogía medido el Tamaño en CFP, en una historia sucesiva de estimaciones.
- H1_e: La media del MRE de la estimación de experto es menor que la media del MRE de la estimación por Analogía medida la Complejidad en P, en una historia sucesiva de estimaciones.

4.1 Estimación de CU de una aplicación

En primer lugar se analiza la aplicación en su conjunto, por lo tanto se consideran todos los CU para comparar los métodos formales de estimación con la estimación del experto.

Métodos formales de estimación.

Se consideraron dos métodos formales: Regresión lineal y Analogía.

Regresión lineal.

Se planteó el modelo de regresión lineal $Y = a + b X$ donde Y es el Esfuerzo Real y X es CFP o P. La Tabla 4 muestra los resultados de la regresión lineal. No se obtuvo un modelo significativo usando como variable independiente CFP, como se puede observar en la Tabla 4 el valor de R^2 Ajustado es muy bajo y p-value es mayor a 0.05. Por el contrario fue posible obtener el valor de la recta de regresión usando como variable independiente a P obteniendo un R^2 Ajustado igual a 0.50 y un p-value igual a 0.001, eliminando dos outliers. También se testeó la condición de normalidad de los residuos aplicando el test de normalidad Shapiro-Wilk obteniendo un p-value igual a 0.21.

Table 4. Método de regresión lineal

EE= a + b X	Outliers	R²	R² Ajustado	p-value
--	--	0.07	0.02	0.26
EE=58.78 +5 * P	CU ₁₂ y CU ₁₆	0.52	0.50	0.001

Para el cálculo de los errores en la estimación se usó el método de “cross-validation”, eliminando del modelo cada caso de uso a estimar. La Tabla 5 muestra los valores de la media (MMRE) y mediana (MeMRE) de los MRE, la calidad de la predicción (PQ) y los ErR.

Table 5. Comparación de los métodos de estimación

Método de Estimación	MMRE	MeMRE	PQ(0.25)	ErR (min..max)
Regresión Lineal (P)	1.45	0.34	0.38	-8.38..0.79
Analogía (CFP)	1.53	0.83	0.06	-6.7..0.87
Analogía (P)	0.94	0.46	0.39	-4.52..0.89
Experto	0.19	0.18	0.72	-0.21..0.5

Analogía.

Este método de estimación consiste en encontrar un proyecto similar al proyecto a estimar basándose en una característica. En forma independiente se uso Tamaño_[CFP] o Complejidad_[P]. Se considera la productividad del CU “más similar” en Tamaño o Complejidad para la obtención del EE. Entonces, para el cálculo de la Productividad (PR) se utiliza la fórmula 4 [19],

$$PR = ER / X \tag{4}$$

Siendo X Tamaño_[CFP] o Complejidad_[P]. Se obtiene el EE utilizando la fórmula 5,

$$EE = X * PR_{PA} \tag{5}$$

Siendo el PR_{PA} la productividad del proyecto análogo, la productividad del proyecto que tiene un valor más cercano en Tamaño_[CFP] o Complejidad_[P]. En caso de existir más de un proyecto análogo se toma el valor promedio de la productividad de los proyectos análogos.

Por ejemplo, CU₁ la Tabla 1 de tiene un esfuerzo de 264 HH y un tamaño de 20 CFP. El CU más cercano en Tamaño es CU₅ con un CFP de 16 y una ER de 104 HH, PR_{CU5} es igual al cociente entre 104 y 16, lo que equivale 6.5. Por lo tanto, EE_{CU1} es igual a 130 HH. La Tabla 5 muestra el análisis estadístico de los valores obtenidos usando CFP o P.

Comparación de los métodos formales con la estimación del experto.

La Tabla 5 compara los métodos de estimación formales con el método de estimación basado en el experto. Usando un método de estimación basado en expertos se obtuvo un resultado aceptado para una técnica de estimación: errores menores a un 25% [18].

En la Tabla 5 se observa que al emplear regresión lineal, solo se obtuvo valores significativos cuando se uso P como variable independiente. Solo el 38% de las estimaciones tuvieron un error menor al 25%, siendo este porcentaje bajo.

Al emplear estimación por analogía los resultados correspondiente a CFP fueron muy bajo, es decir, solo el 6% de las estimaciones realizadas por analogía tomando como variable dependiente a CFP pudieron tener un error menor al 25%. Por el contrario al emplear P como variable independiente se pudo mejorar el resultado: el 39% de las estimaciones tuvieron un error menor al 25%.

Si comparamos ambas técnicas podemos concluir que usar P como medida de Complejidad es mejor que usar CFP como medida de Tamaño. Al mismo tiempo, ninguna de estas técnicas para este caso de estudio es mejor que la estimación dada por el experto. Como se observa el 72% de las estimaciones del experto tiene un error menor al 25%.

Se puede concluir entonces que ninguna de las técnicas de estimación tradicionales pudo superar la estimación del experto. Existen algunas que son mejores que otras pero no son equiparables a la experiencia del experto.

4.2 Estimación sucesivas de versiones de una aplicación

En un escenario en el cual se estiman sucesivamente versiones de una aplicación se espera que el experto aprenda de las estimaciones anteriores. La Tabla 1 muestra los CU agrupados por versiones. Se compara la estimación del experto con un método formal basado en analogías.

El método de estimación por analogía usando como variables a CFP y P se aplico en la misma forma que se explicó anteriormente, con la particularidad que el conjunto de CU usados para encontrar el análogo, es el subconjunto de los CU de la versión correspondiente más los CU de las versiones anteriores. Por ejemplo, cuando se estima la versión 10 se usan los CU de la versión 7 y 9. Dada esta limitación en el cálculo no fue posible aplicar regresión lineal puesto que en algunos casos la cantidad de puntos a considerar en el modelo era muy pequeña, obteniendo modelos no significativos.

Para obtener los resultados del experto, se realizo una segunda entrevista. Pero esta vez se le mostro solo los CU a estimar por versión y los errores cometidos al estimar las versiones anteriores. Por ejemplo, al estimar la versión 9, se le mostró los valores reales de la versión 7, además de la estimación realizada anteriormente de esta versión. La Tabla 1 muestra las estimaciones sucesivas realizadas por el experto. La Tabla 6 muestra los errores de las estimaciones sucesivas, aplicando el método por Analogías (CFP y P) y basado en expertos.

Como se ve en la tabla 6 las MMRE de la estimación del experto tiene un rango de [0.18-0.22], y los PQ [0.63-0.75], valores mejores que los restantes. En el caso de Paths el rango del MMRE es de [0.76-2.22] y el del PQ es de [0-0.39]. Los valores obtenidos usando COSMIC son MMRE [1.18-1.53] y PQ [0-0.25]. Si bien es mejor el resultado obtenido usando la métrica Paths y el método de estimación por Analogía no hay una mejora significativa si se lo compara con COSMIC.

Las estimaciones sucesivas realizadas por el experto fluctúan pero la media y mediana de MRE se mantienen dentro de valores aceptables para la industria, esto es menores al 25% y similares a la estimación del experto mostrada en Tabla 5. Los valores de PQ son los más altos de la Tabla 6 y si consideramos el PQ de la primera medición (72%), podemos concluir que la información más detallada no ha mejorado la precisión del experto. El experto al conocer su error introduce en la corrección de este una desviación mayor con respecto al valor real. Al mismo tiempo se observa en la Tabla 6 un aprendizaje por parte del experto, no logrando un valor mejor por las limitaciones de la mente humana.

Las estimaciones realizadas por Analogía y la métrica Paths también fluctúan, pero el PQ tiende a mejorar. Usando la métrica COSMIC también varía, reportando el PQ una mejora no sensible. En ambos casos el hecho de contar con una cantidad de CU menores para seleccionar el más análogo no mejora el error en la estimación.

Table 6. Comparación de la estimación sucesiva de versiones

Método de Estimación	MMRE	MeMRE	PQ(0,25)	Errores relativos (min..max)
COSMIC v7	1.35	0.92	0.25	-3.36.. 0.2
COSMIC v9	1.18	0.9	0	-3.36.. 0.5
COSMIC v10	1.43	0.75	0	-6.7..0.87
COSMIC v11	1.29	0.79	0	-6.7..0.87
COSMIC v12	1.53	0.83	0.06	-6.7..0.87
Paths v7	2.22	0.66	0	-7.25..0.88
Paths v9	1.86	0.95	0	-7.25..0.88
Paths v10	1.01	0.83	0.1	-4..0.83
Paths v11	0.76	0.46	0.33	-4..0.83
Paths v12	0.94	0.46	0.39	-4.52..0.89
Experto (V7)	0.18	0.19	0.75	0.09..0.25
Experto (V9)	0.20	0.22	0.67	0.09..0.29
Experto (V10)	0.26	0.21	0.64	-1..0.35
Experto (V11)	0.22	0.17	0.67	-1..0.35
Experto (V12)	0.20	0.17	0.63	-1..0.35

4.3 Amenazas de validez

La mayor amenaza de validez de este caso de estudio es la registración del ER realizado por la empresa. Es conocido que estos registros no son exactos, por lo que fueron validados por registros alternativos de horas trabajadas. Concluyendo que son registros correctos.

Puede llamar la atención la omisión de la versión 8. Esta fue descartada por falta de calidad de la descripción textual de los casos de usos.

También la cantidad de casos de usos utilizados es limitada y podría afectar los resultados. En la selección de los casos de uso se selecciono un período de versiones que tuvieran registrado y disponible el ER. El tiempo de medición fue otra variable que fue necesario tener en cuenta considerando que era una aplicación compleja. Si

bien la cantidad de datos es acotada la comprobación estadística de la hipótesis planteada es significativa.

Otro aspecto que puede influir es el conocimiento que tiene el experto de las horas reales de este caso de estudio. El experto se incorporó a la empresa en una etapa posterior al desarrollo real de las versiones presentadas en este caso de estudio y no tenía conocimiento de las horas reales. Además, el grupo que desarrollo las versiones incluidas ha variado con respecto al momento en que se consultó al experto.

4.4 Conclusiones del caso de estudio

Para realizar la verificación estadística se utiliza el test no-paramétrico Wilcoxon dado que las distribuciones no son normales. Se selecciona este test, puesto que para pruebas pareadas es posible aplicarlo a distribuciones continuas. Fue posible rechazar la hipótesis nula a favor de la hipótesis alternativa $H1_a$ aplicando el test no-paramétrico Wilcoxon p-value igual a 0.009. También fue posible rechazar la hipótesis nula a favor de la hipótesis alternativa $H1_b$ aplicando el test no-paramétrico Wilcoxon p-value igual a 0.000, $H1_c$ p-value igual a 0.006, $H1_d$ p-value igual a 0.000 y $H1_e$ p-value igual a 0.000.

Por lo tanto se concluye que en este caso de estudio NO fue posible superar con las estimaciones de expertos utilizando métodos de estimación formales (Regresión lineal y Analogía), tanto en una estimación para todos los CU de la aplicación como para una estimación sucesiva de versiones.

La participación de un experto limita la conclusión del caso de estudio, pero no la descarta, destacando el valor de todo caso de estudio en un ámbito real. Al mismo tiempo es importante comprender que el experto fue tipificado, esperando obtener diferentes precisiones al variar el perfil del experto [20].

5 Conclusión final

Como ha sido anticipado por Jorgensen [1] existen situaciones donde se puede esperar que las estimaciones expertas sean más precisas que los métodos formales de estimación. En el caso de estudio presentado el experto tiene una alta experiencia en el desarrollo de software, un alto conocimiento de la tecnología y un conocimiento medio en el dominio y en el rendimiento del grupo. Estos aspectos favorecieron la precisión de sus estimaciones.

Sorprendió que en la historia sucesiva de estimaciones no se obtuviera una mejora en la precisión. Se cree que esto se debe a la imprecisión de los ajustes humanos, aunque se muestra un aprendizaje en la sucesión de estimaciones, logrando en la estimación final un valor similar obtenido en la estimación de todos los casos de uso de la aplicación.

En el caso de los métodos formales de estimación, se usaron modelos de una sola variable: Tamaño o Complejidad. Esto afecta la precisión de los modelos dado que si bien estos factores son los más importantes, sus variaciones no explican en un alto porcentaje la variación del esfuerzo.

El presente trabajo aporta un caso de estudio usando métricas no relevadas en Jorgensen [1] y pone en evidencia la obtención de unos resultados particulares destacando las características del experto en un proyecto complejo del ámbito industrial. Para poder generalizar sus resultados es necesario analizar otros productos de diversos dominios, incorporar expertos con perfiles diferentes y otros métodos formales.

Sería interesante en trabajos futuros trabajar con modelos que incluyan otras variables que afecten la estimación de esfuerzo y variar el perfil de los expertos, particularmente observar la variación de la precisión del experto al tener un conocimiento del rendimiento de los perfiles del grupo de desarrollo y del dominio alto.

Agradecimientos. El presente proyecto se ha realizado con el apoyo de la Universidad Austral y la Universidad Argentina de la Empresa.

Referencias

1. Jorgensen, M.: A review of studies on expert estimation of software development effort. *Journal on System and Software*, Vol. 70, No. 1-2, 37-60 (2004).
2. Jorgensen, M. y Shepperd.. A systematic review of software development cost estimation studies, *IEEE Transactions on Software Engineering*, Vol. 33, No. 1. p. 3-53, (2007).
3. Montgomery, D, Peck, E.A., Vining, G.G.: *Introducción al análisis de regresión Lineal*, Compañía Editorial Continental (2004)
4. Shepperd, M., Schofield, C.: Estimating Software Project Effort Using Analogies. *IEEE Trans. on Software Eng.*, vol. 23, no. 11, pp. 736-743, Nov. (1997).
5. COSMIC – Common Software Measurement International Consortium, 2009: *The COSMIC Functional Size Measurement Method - version 3.0.1 Measurement Manual (The COSMIC Implementation Guide for ISO/IEC 19761: 2003)*, May (2009).
6. COCOMO II Model Definition Manual. http://csse.usc.edu/csse/research/COCOMOII/cocomo_downloads.htm
7. Robiolo, G., Badano, C., Orosco, R.: Transactions and Paths: two use case based metrics which improve the early effort estimation. *ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '09)*, Lake Buena Vista, FL, October (2009).
8. Lavazza, L., Robiolo, G.: Introducing the Evaluation of Complexity in Functional Size Measurement: a UML-based Approach. *Symposium on Empirical Software Engineering and Measurement (ESEM)*, Boszano-Bozen, Italia, Sept 16-17 (2010).
9. McCabe, T. A: Complexity measure, *IEEE Transactions on software Engineering*, Vol. SE-2, NO. 4. (1976)
10. Jorgensen, M and Gruschke, Tanja M.: The role of outcome feedback in improving the uncertainty assessment of software development effort estimates. *ACM Trans. Softw. Eng. Methodol.* 17, 4, Article 20 (August 2008), 35 pages. (2008)
11. Jorgensen, M. and Halkjelsvik, T: The effects of request formats on judgment-based effort estimation, *Journal of Systems and Software*, 83 (1), 29-36. (2010)

12. Mendes, E.: Improving software effort estimation using an expert-centred approach. In Proceedings of the 4th international conference on Human-Centered Software Engineering (HCSE'12), Winckler, M., Forbrig, P. and Bernhaupt R.(Eds.). Springer-Verlag, Berlin, Heidelberg, 18-33 (2012).
13. Cuadrado-Gallego, J.J., Rodríguez-Soria, P. and Martín-Herrera B.: Analogies and Differences between Machine Learning and Expert Based Software Project Effort Estimation. In Proceedings of the 2010 11th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD '10). IEEE Computer Society, Washington, DC, USA, 269-276 (2010).
14. Jørgensen, M.: Forecasting of software development work effort: Evidence on expert judgement and formal models , *International Journal of Forecasting* 23 449–462 (2007)
15. Jorgensen, M. and Boehm, B.: Software Development Effort Estimation: Formal Models or Expert Judgement? *IEEE Software* (March/April):14-19 (2008).
16. Hughes, R.T.: Expert judgement as an estimating method, *Information and Software Technology*, Volume 38, Issue 2, 67-75, (1996).
17. MacDonell, S. G., & Shepperd, M. J.: Combining techniques to optimize effort predictions in software project management. *Journal of Systems and Software*, 66(2), 91–98 (2003).
18. Fenton, N.E. and Pfleeger, S.L.: *Software Metrics*. PWS Publishing Company (1997)
19. Jørgensen, M., Indahl, U and Sjøberg, D.: Software Effort Estimation by Analogy and "Regression Toward the Mean". *Journal of Systems and Software* 68(3): 253-262 (2003)
20. Halstead, S., Ortiz, R., Córdova, M. and Seguí, M.: The impact of lack in domain or technology experience on the accuracy of expert effort estimates in software projects. In Proceedings of the 13th international conference on Product-Focused Software Process Improvement (PROFES'12), Dieste, O., Jedlitschka, A. and Juristo, N. (Eds.). Springer-Verlag, Berlin, Heidelberg, 248-259, (2012).