

ANEXO TÉCNICO

Se presentan a continuación algunas consideraciones respecto del estudio de *“Evaluación en profundidad del Programa Franquicia Tributaria para capacitación”*.

En primer lugar, nos parece importante destacar –tal cual lo hace el informe final- que en este estudio se ha empleado una batería de instrumentos e indicadores, con el propósito de que el impacto del Programa F T se observe desde varias miradas complementarias y su comprensión se torne integral. Es decir, más allá del resultado de una cifra, lo que importa es relacionar el conjunto de indicadores que brotan desde los distintos instrumentos aplicados y entender cada uno de ellos a partir de su singular naturaleza y lo que específicamente está diciendo. De esta manera, cobran sentido y aportan al conjunto desde su particular perspectiva.

En lo que sigue, se recomienda que el lector se encuentre familiarizado con los rudimentos básicos de la estadística y del cálculo diferencial, al mismo tiempo que esté totalmente compenetrado con los términos de referencia de este estudio, la propuesta técnica y los acuerdos que se fueron alcanzando con el SENCE en el transcurso del trabajo.

Las consideraciones dicen relación con los siguientes puntos:

1. Configuración del grupo de estudio y del grupo de control y su comparación
 2. El tratamiento que se aplicó al método de regresión econométrica que establece la DIPRES
 3. Sobre el PSM se aclaran los siguientes puntos:
 - a. El pseudo R^2
 - b. Los intervalos de confianza
 - c. Los regresores en el Logit.
 4. Tamaño Muestral
-
1. Sobre la configuración de ambos grupos, nos parece necesario acotar lo siguiente:
 - a. La selección del grupo de beneficiarios y de control fue realizada en empresas que participaron en el programa de Franquicia tributaria durante el período (2002 – 2007). En estas empresas, se trabajó junto a los encargados de RRHH, una selección aleatoria (sobre las bases de datos del SENCE y de las empresas) de los trabajadores, poniendo especial énfasis en que los trabajadores que no tuvieron capacitación vía Franquicia Tributaria durante el mismo período de estudio

(controles) fueran lo más similares posible a aquellos que sí fueron capacitados en el período.

b. Comparación de grupo estudio y grupo control:

Edad Promedio (años)

Grupo	Edad
Control	38
Estudio	40

Años de escolaridad promedio

Grupo	Años de escolaridad
Control	13,92
Estudio	13,94

Nivel de Estudios

Nivel	Control	Estudio
<i>Sin estudios</i>	0,0%	0,4%
<i>Básica incompleta</i>	1,4%	1,1%
<i>Básica completa</i>	4,1%	1,9%
<i>Media CH incompleta</i>	6,0%	4,5%
<i>Media TP incompleta</i>	2,6%	3,7%
<i>Media CH completa</i>	10,8%	12,4%
<i>Media TP completa</i>	14,4%	16,0%
<i>CFT incompleta</i>	5,3%	6,8%
<i>CFT completa</i>	10,6%	8,7%
<i>IP incompleta</i>	4,8%	4,6%
<i>IP completa</i>	10,1%	8,2%
<i>Universitaria incompleta</i>	7,9%	7,4%
<i>Universitaria completa</i>	21,9%	24,2%
<i>Total General</i>	100,0%	100,0%

Nivel ocupacional

Grupo	Sin Jefatura	Jefatura Intermedia	Jefatura Máxima
Control	73,1%	19,6%	7,4%
Estudio	73,0%	23,2%	3,8%

Situación contractual

Grupo	Honorarios	Contrato Definido	Contrato Indefinido
Control	3,8%	13,8%	82,4%
Estudio	1,0%	16,4%	82,6%

Jornada Laboral (horas semanales)

Grupo	40	45	48
Control	4,6%	85,0%	10,4%
Estudio	2,8%	88,6%	8,5%

Como se desprende de inmediato, ambos grupos no son sólo parecidos, sino prácticamente idénticos. Esto se ve posteriormente refrendado por los resultados que arroja el examen comparativo de los capitales intangibles y la empleabilidad y, los resultados del Logit y el posterior pareo que comprende el PSM.

2. Medición de Capitales Intangibles

Con relación a los resultados principales de la medición de capitales intangibles, se consigna el siguiente cuadro general y sus comentarios asociados (pág. 155-156 del informe final):

Grupo de Estudio

	Estudios	Capital Humano	Capital Simbólico	Capital Cultural	Capital Social	Capital Psico Social	Capital Cognitivo	Empleabilidad
N	734	734	734	734	734	734	734	734
Missing	0	0	0	0	0	0	0	0
Media	653	699	655	745	644	795	490	659
Mediana	591	660	671	754	691	810	460	667
Desviación Estándar	128	93	61	43	128	50	200	62
Varianza	16.280	8.694	3.725	1.813	16.273	2.510	40.005	3.876
Skewness	0	0	-2	-2	-2	-5	0	-1

Kurtosis	0	0	6	7	3	24	-1	0
Rango	664	465	415	371	703	368	720	334
Mínimo	156	355	321	435	100	452	100	443
Máximo	820	820	736	805	803	820	820	777

Grupo de Control

	Estudios	Capital Humano	Capital Simbólico	Capital Cultural	Capital Social	Capital Psico Social	Capital Cognitivo	Empleabilidad
N	416	418	418	418	418	418	418	418
Missing	2	0	0	0	0	0	0	0
Media	652	690	661	747	653	794	467	656
Mediana	591	660	680	755	710	809	460	664
Desviación Estándar	127	99	62	42	133	49	225	69
Varianza	16.128	9.773	3.857	1.733	17.784	2.392	50.607	4.705
Skewness	0	-1	-2	-2	-2	-4	0	-1
Kurtosis	-1	1	7	6	3	24	-1	1
Rango	570	574	415	317	697	393	720	409
Mínimo	250	246	323	490	100	427	100	377
Máximo	820	820	738	807	797	820	820	787

La empleabilidad entre ambas poblaciones -si consideramos la escala- es idéntica: 659 y 656, arrojando una variación de solamente un 0,6%.

Ambas poblaciones ostentan capitales prácticamente idénticos; las variaciones en cada uno alcanzan hasta un máximo de un 1%, exceptuando el capital cognitivo que llega a un 5%. Estas condiciones están expresando la consistencia entre ambos grupos e igualdad entre estudios y controles.

Resalta especialmente la equivalencia de los capitales humanos (1% de diferencia), atributo clave para la determinación de la homogeneidad entre ambas poblaciones. Si los capitales humanos, en su dimensión intangible, son idénticos, entonces las variaciones que se registren en su expresión tangible (ingresos) son decidoras.

En particular, es notable la identidad plena en la categoría de estudios, la cual marca una diferencia de 0,2%. Esto es crucial, puesto que una igualdad absoluta en el nivel de estudios da cuenta de dos poblaciones cuya base esencial es la misma para la medición de los impactos, los cuales se manifiestan a través de las diferencias entre los ingresos en el tiempo.

Las diferencias en los niveles de estudio en las personas se reflejan en diferencias expresivas de ingresos en el mundo laboral. Es por ello que la identidad que se manifiesta en este aspecto, entre las poblaciones de estudio y control, adquiere una relevancia central para el cálculo de impacto del Programa de Franquicia Tributaria.

Puede afirmarse, con toda propiedad, que ambas poblaciones son idénticas con relación a la empleabilidad en general y a cada capital intangible en particular. No se advierten sesgos en variables no observables.

Se está en presencia, por lo tanto, de un pareo adecuado entre ambas poblaciones, cuestión clave para un estudio de impacto. Esto mismo se corrobora más adelante con el *matching* que se efectúa mediante el método del *Propensity Score Matching*.

En consecuencia, los resultados que surjan desde la evaluación de impacto van a medir exactamente el aporte neto del Programa en los incrementos de capital humano expresados en ingresos. Es decir, frente a atributos idénticos –en lo intangible- serán los ingresos los que reportarán las diferencias que se manifiestan en los capitales humanos –en su dimensión tangible- de ambos grupos, lo cual tributará específicamente al valor de la capacitación debido a la Franquicia Tributaria.

Esto es de crucial relevancia, puesto que se está garantizando que la propensión a ser tratado es independiente de las características de los grupos de control y tratados, y como se dijo anteriormente, se captura el tratamiento por las diferencias de ingresos.

Los resultados del PSM, en lo referido a este aspecto, forman parte del punto 4 del presente anexo.

3. Regresiones:

- a. Lo principal en la utilización de esta técnica es que el modelo de regresores (o variables independientes) ocupado, pertenece a la autoría de expertos internacionales en esta materia: Deheja y Wahba (1999) y que se citan en la página 178 del informe final.

No obstante, en el trabajo se aplican varios test, cuyo desarrollo y resultados se consignan en anexos y ellos arrojan que el modelo **tiene variables omitidas y que no es lineal**, razón por la cual indicamos expresamente que este modelo no es robusto.

Es por ello que se señala en el informe que no se puede seguir probando con los datos de una sola muestra, bajo un esquema que **fuerza relaciones lineales** en las variables.

El desafío consiste justamente en la determinación de un modelo que verdaderamente represente el fenómeno en estudio, el cual es materia de una investigación profunda y constituye un estudio en sí mismo y que, a su vez, lleva tiempo diseñarlo y hay que trabajarlo mediante el análisis de múltiples muestras de un mismo universo y, también, longitudinalmente.

Por esto es que falla el modelo de regresión establecido en los términos de referencia para cumplir con las especificaciones de la DIPRES; y así se deja consignado en el informe final, donde **se privilegia explícitamente el método del PSM** (que también lo sugiere la DIPRES), el cual sí es significativo estadísticamente y, carece de las anteriores fallas y, por lo tanto como se hace internacionalmente, se recogen sus resultados como más confiables.

- b. En torno a la endogeneidad de variables, lo que realmente ocurre es lo siguiente: internacionalmente se reconoce que habría cierta causalidad doble entre las

variables educación e ingreso, en el sentido de que mayor ingreso causa mayor educación y viceversa. Por lo tanto, al modelo en general habría que hacerle una corrección a la Heckman: utilizando variables instrumentales y hacer una estimación de dos etapas.

El problema que está presente en todos los estudios internacionales –y se reconoce como tal- es encontrar los instrumentos adecuados para las variables endógenas. Este problema nadie lo corrige porque encontrar una variable instrumental para la resolución de este problema, puede significar un trabajo de años.

Dicho de manera un poco más simple, encontrar una variable que esté correlacionada con la educación (por ejemplo) y que a su vez no esté correlacionada con el ingreso, es muy difícil. Otro ejemplo de esto es utilizar una variable de coeficiente de intelectual en vez de educación, pero nuevamente no se sabe qué causa qué ¿la educación hace que suba el CI? ¿o el CI causa más educación? Dado esto, hay que encontrar variables instrumentales que no se conocen y que son muy difíciles de definir, ya que estamos hablando de variables blandas o cualitativas. Por ello, es mejor utilizar las variables que intuitivamente sabemos que afectan al ingreso y preocuparnos de interpretar bien el resultado del parámetro de cada variable, puesto que corregir este problema se transforma en un estudio propio y muy complejo en sí.

Como esto es reconocido a nivel internacional, en todo el mundo se procura solucionarlo con el PSM, el cual no adolece de este defecto y es reconocido entonces como mucho más robusto y certero.

No obstante, independientemente de cuál sea el modelo de regresión adecuado y sus variables asociadas, hay que rescatar (y se menciona en el informe) el resultado de impacto. Lo que finalmente interesa de este método es el impacto, el cual es representado por el coeficiente beta asociado a la variable de interacción ($G*T$). Y, como se señala en el informe, aún sin significancia estadística se sitúa en los órdenes del resto de los indicadores, razón por la cual se mantuvo.

Con todo, el compromiso metodológico –establecido en la propuesta técnica- alude expresamente a una regresión que se explicita claramente. No sostuvimos en ninguna parte que nos íbamos a hacer cargo de las insuficiencias de uno u otro método (que todos las tienen) y realizar exhaustivos estudios explorando todos los caminos metodológicos posibles (que son múltiples), pues ello escapa a la labor de esta consultoría y pasa a formar parte de una investigación de ciencia aplicada, tal como ya se señaló.

4. En relación con el PSM, queremos señalar lo siguiente:

- a. En primer lugar, el pseudo R^2 es bajo (0.06) y ello no sólo no es erróneo, sino que está señalando de antemano la plena identidad entre el grupo de estudio y el grupo de control. Es decir, está reafirmando lo ya señalado y confirmado anteriormente.

Las cifras matemáticas siempre deben asentarse en la racionalidad del fenómeno en estudio, en un sustrato de realidad. Por esta razón es que el pseudo R^2 no siempre debe ser alto; hay que detenerse a reflexionar acerca de qué es lo que se está haciendo y midiendo, en el caso del PSM, con una regresión Logit.

Precisamente en este caso, mientras más bajo es el pseudo R^2 , mejor es la semejanza entre los grupos de estudio y control, debido a que es la probabilidad de ser tratado de lo que está dando cuenta el Logit y para ello se contrasta a ambos grupos en relación con determinadas características individuales que deben ser comunes a ambos (soporte común). Y por ello, si las variables tienen bajo poder explicativo es porque no discriminan entre ambos grupos, exactamente porque éstos son muy parecidos. Y, como ya se anotó en el acápite de los capitales intangibles: “la propensión a ser tratado es independiente de las características de los grupos de control y tratados, y como se dijo anteriormente, se captura el tratamiento por las diferencias de ingresos”.

Es más, los expertos internacionales recomiendan que se pruebe de nuevo el Propensity (Logit) también después del pareo, con el objeto de cerciorarse que precisamente el pseudo R^2 sea bajo y que el test F no arroje significancia estadística, es decir, cerciorarse de que ambos grupos (tratados y controles) quedaron completamente pareados y que son prácticamente iguales. Al respecto se consigna la siguiente cita de Caliendo y Kopeinig:

“IZA DP No. 1588

Some Practical Guidance for the Implementation of Propensity Score Matching

Marco Caliendo

Sabine Kopeinig

May 2005

Joint significance and Pseudo- R^2 : Additionally, Sianesi (2004) suggests to re-estimate the propensity score on the matched sample, that is only on participants and matched non-participants and compare the pseudo- R^2 's before and after matching. The pseudo- R^2 indicates how well the regressors X explain the participation probability. After matching there should be no systematic differences in the distribution of covariates between both groups and therefore, the pseudo- R^2 should be fairly low.”

El PSM es utilizado sobre la base de que existe diferencia entre estudios y controles en un estudio cuasi experimental. Para ello se concibió y con el objeto de

igualarlos en las variables que les son comúnmente características, mediante el pareo sobre sus distancias en torno a las asignaciones de probabilidad de ser tratado.

En consecuencia, si el pseudo R^2 resulta bajo antes de que se efectúe el pareo, lo que esta cifra está señalando con toda claridad y evidencia, es que ambos grupos (estudio y controles) son prácticamente idénticos: situación ideal y que reafirma la correcta selección de la muestra del grupo de control.

- b. En torno a la supuesta ausencia de los intervalos de confianza, reproducimos las tablas que están en el informe en páginas 175 y 177:

logit regression						Number of obs	1135
						=	
						LR chi2(8)	84.39
						=	
						Prob > chi2	0
						=	
Log likelihood = -696.153						Pseudo R2	0.06
						=	
Variable	Coeficiente	Errores Estándar	Estadístico Z	P>z	[95% Conf. Interval]		
Edad	0.283**	0.041	6.930	0.000	0.203	0.363	
Años de escolaridad	-0.063	0.073	-0.870	0.382	-0.206	0.079	
Estudios	0.180	0.163	1.110	0.267	-0.138	0.499	
Género	0.315***	0.138	2.280	0.022	0.044	0.585	
Estado civil	-0.002	0.007	-0.270	0.790	-0.016	0.012	
Edad al cuadrado	0.003***	0.000	-6.640	0.000	-0.004	-0.002	
Estudios al cuadrado	-0.008	0.008	-1.030	0.304	-0.023	0.007	
Años de esc. al cuadrado	0.001	0.001	0.830	0.408	-0.001	0.002	
Salario ex-ante	0.000***	0.000	3.050	0.002	0.000	0.000	
Salario ex-ante al cuadrado	0.000***	0.000	-2.310	0.021	0.000	0.000	
Constante	-5.474	0.959	-5.990	0.000	-7.627	-3.867	

Errores Estándares Robustos in paréntesis calados con Bootstrapping

Cada asterisco da la significancia a *** $p < 1\%$, ** $p < 5\%$, * $p < 10\%$

Los intervalos de confianza se exhiben en las dos últimas columnas. Es recomendable cotejarlos con los valores del p-value y en relación con la variable z normalizada. Cada p-value da cuenta de la significancia estadística exacta y cada z (o t normalizado) corresponde a una probabilidad específica en la distribución de probabilidad normal (gaussiana). Más aún, en la tabla al final se explicita que los asteriscos equivalen a las distintas significancias estadísticas.

Igualmente ocurre con el resultado final:

No de Tratados.	No de Controles	ATT	Error Estándar	<i>Estadístico t</i>
732	400	24870.09	6821.123	3.646

El estadístico t de la distribución de Student equivale a una significancia estadística al 1% (la mejor, óptima). Es decir, el impacto ATT (\$24.870) es estadísticamente significativo al 1%. Esto quiere decir que se rechaza la hipótesis nula de que el coeficiente es cero; el modelo explica muy bien la variable salarios y , por lo tanto, el impacto de diferencias en diferencias del PSM (ATT) es válido y totalmente significativo estadísticamente (cuestión que no ocurre con el modelo de regresión lineal y , por ello, es que se reitera que este modelo del PSM es el más robusto y atendible en sus resultados).

Nuevamente aquí, al igual que en el caso de la variable normalizada z (o t normalizado), cada t corresponde a una probabilidad dentro de la distribución de probabilidad de Student.

- c. Acerca de la ausencia de regresores en el PSM.

Este modelo corresponde a Deheja y Wahba (1999), ya citados.

Toda la literatura experta internacional recomienda que el vector de regresores (*covariates*) tenga baja dimensionalidad, es decir, que tenga el mínimo de regresores, pues, tal como se explica en el informe final en la página 175:

“Segundo, se podría cuestionar el pareo por utilizar una o demasiadas variables características, sin embargo, el problema de incluir más o menos variables tiene un costo de oportunidad en cuanto a Varianza y Consistencia del estimador de la diferencia de los tratados y no tratados. Bryson, Dorsett y Purdon (2002) muestran por qué se debe evitar la sobreparametrización del modelo, es decir, incluir más variables características, aunque el PSM no quede balanceado. Ellos recomiendan usar las variables que permiten un balanceo perfecto del PSM, porque al incluir más se exagera la varianza, por ende, uno podría llegar a resultados erróneos en cuanto a significancia estadística en el parámetro de interés.”

¿Qué quiere decir esto? Que mientras más variables se incluyan, aumenta la varianza y se pierde precisión en la variable dependiente (asignación de probabilidad de ser tratado). Al perder precisión esta variable clave, se invalida toda la métrica posterior destinada a medir distancias de los controles con los tratados. Y por lo tanto, el pareo (que es la meta final del PSM) termina diciendo nada y careciendo de sentido.

Lo complementamos con la siguiente cita:

Microeconometrics

Cameron y Trevedi

CHAPTER 25

Treatment Evaluation

The procedure is to match treated and untreated individuals on the inobservable characteristics x .

Exact matching is practicable when the vector of covariates is discrete and the Sample contains many observations at each distinct value of x_i . If the covariate vector has a high dimension, or if continuous variations among some Covariates are present, then exact matching between treated and non treated groups becomes impractical. This problem motivates inexact matching methods.”

Lo anterior también alude al **soporte común**. El soporte común consiste en un vector de regresores que verdaderamente discrimine entre tratados y controles mediante variables que son comunes a ambos. Para eso se aplica el algoritmo de balanceo de Becker e Ichino (consignado en el informe). Este algoritmo tiene la ventaja de realizar la asignación de probabilidad según las variables de características y a la vez probar que la propiedad de “Balance” se cumpla; esto quiere decir que los individuos que son tratados tengan las mismas características

y así garantizar que tengan la misma probabilidad de ser elegidos para el Programa. (página 174 del informe final).

Éstos fueron sometidos al algoritmo de asignación de probabilidad que, a su vez iba probando el “balanceo” de las características entre los tratados, es decir, el mismo número de vectores característicos para los tratados. Esta probabilidad fue estimada a través de un modelo Logit y luego se hizo la prueba de balanceo para las variables características y todas pasan el test.

Ahora bien, aún cuando se hubiera pensado en agregar otras variables, las cuales tendrían que estar asociadas a las condiciones laborales, como por ejemplo la jornada laboral, situación contractual, nivel de responsabilidad, se habría incurrido en lo siguiente:

Primer error: alta dimensionalidad del vector de regresores, con las consecuencias mencionadas de incremento de varianza y alteración de confiabilidad: pérdida de precisión en la asignación de probabilidad y métrica dudosa, lo cual lleva a un pareo incierto.

Segundo, hay que tener cuidado en ocupar como regresores variables discretas (o cualitativas) multinivel, pues éstas carecen de varianza y no sirven como variables independientes (obstruyen una regresión).

Tercero, las variables (regresores) deben tener poder explicativo o discriminatorio. Por lo tanto, aquellas variables en que los estudios y controles son iguales carece de sentido incluirlas, como lo demuestran los cuadros comparativos entre estudios y controles, consignados al principio y que aluden a estas variables laborales mencionadas.

Es en este sentido que se apunta a que las cifras matemáticas deben tener un sustrato en la realidad, que le otorga racionalidad y, por ende, hay que estudiar las variables una por una en su naturaleza, para averiguar si tiene o no sentido incorporarlas en el vector de regresores.

En virtud de ello es que se descartó su inclusión en este caso, dada la plena identidad entre estudios y controles que, finalmente, es lo principal.

De este modo, el PSM pretende resolver la existencia de disparidad en variables observables y si ella no existe, obviamente es mejor, en función de la optimización de los resultados.

Las características inobservables (como motivación, fortaleza identitaria, confianza en sí mismo, redes sociales, etc, etc.) las zanja el método de capitales intangibles.

5. Tamaño muestral:

- a. Desde el principio sostuvimos ante el SENCE que este estudio debía hacerse con muestras autorrepresentativas por año.

El punto es que, tal como lo conversamos en más de una oportunidad con el SENCE, los recursos impedían hacerlo.

Con todo, lo central en este aspecto es atenerse a las bases técnicas y administrativas del estudio, junto a la propuesta técnica ofertada.

- b. También sostuvimos que ese hecho, si bien no invalida los cálculos por estratificación, carece de la representatividad estadística que confiere una muestra autorrepresentativa por cada año; no obstante, fue el propio SENCE quien insistió en que igualmente efectuáramos los cálculos de impacto por año.
- c. A pesar de ello, los resultados son muy consistentes: basta leer las cifras. Al final de este anexo se adjunta el consolidado de impactos mediante las distintas técnicas y se comenta el significado de cada uno de los indicadores, su consistencia y se hacen sugerencias en torno a necesaria generación de modelos especiales de impacto, que sean específicos para el Programa F T.
- d. Es por todo ello que se enfatiza en el informe, y se reitera más de una vez, que el modelo del PSM general es el más robusto y atendible en sus resultados.
- e. De otro lado, el compromiso adquirido en la propuesta era alrededor de 420 individuos por cada muestra, entre estudios y controles. Los tamaños muestrales finales fueron de 734 casos de estudios y 419 controles, asumiendo nosotros el costo adicional, aún por encima de los compromisos inicialmente adquiridos.

Con todo, no estamos en presencia de un n bajo. De hecho, tal como se consigna en el informe con este tamaño muestral, el error para el grupo de estudio es de un 3,6%. Y para el grupo de control es de un 4,8%.

6. En torno a la utilización de las regresiones, es menester leer los términos de referencia y parte de los argumentos consignados en el informe y en este mismo anexo.

- a. En primer lugar, en la página número 181 del informe se dice lo siguiente con respecto a las regresiones econométricas:

“... Finalmente y conforme a los resultados que dieron los distintos test, se concluye que este modelo tiene variables omitidas por relaciones no lineales, lo que indica que faltan variables que explican a la variable relevante. Por lo que el resultado es un poco espurio, cuestión que se da típicamente cuando se trabaja con variables categóricas que interactúan entre sí. Por lo tanto, esta metodología

no es determinante para establecer un nivel de impacto, ya que por construcción va a entregar resultados que no son muy finos en su poder explicativo, debido a la imposición de una forma funcional, que fuerza a las variables de manera lineal.”

Al respecto se cita a:

Econometría

Damodar Gujarati

“...son frecuentes los casos en los cuales la teoría no indica la forma funcional precisa del modelo; tampoco dirá si el modelo es lineal en las variables o lineal en los logaritmos de las variables, o alguna mezcla de las dos posibilidades, o si tendrá alguna otra forma”

“Para llegar al modelo adecuado, se podría seguir probando y testeando la transformación de las variables explicativas de manera que reflejen una relación distinta a la lineal, pero dicha tarea puede resultar extenuante debido a que se deja todo en manos de pruebas por ensayo y error. Esto tampoco garantiza que los resultados cambien en el sentido contrario.

Por lo anterior, un modelo que elimine la dependencia de buscar la forma funcional adecuada entre la variable dependiente y las independientes, arrojará resultados mucho más robustos...”

Cita:

Testing Hypothesis of Fitting Models? Another Look at Mass Extinctions”, en Matthew H. Nitecki y Antoni Hoffman, eds., Neutral Models in Biology, Oxford University Press, Oxford, 1987, p. 148, en Gujarati, p. 113

“Cuidado con el chequeo de muchas hipótesis: entre más se torturen los datos, más probable es que ellos confiesen, pero la confesión obtenida bajo presión puede no ser admisible en la corte de la opinión científica”

Estas regresiones fueron aplicadas única y exclusivamente porque así lo exigen los términos de referencia, afirmándose en las especificaciones de la DIPRES. En ningún caso, y tal como lo expresamos, porque sus resultados los consideremos totalmente adecuados y válidos, en virtud de las razones que se esgrimen en el informe.

Como se aprecia, en el mismo informe se declara este método como poco adecuado para obtener impacto. En materia del método, nos atuvimos estrictamente a los términos de referencia y a la oferta presentada en nuestra propuesta técnica.

Terminamos este punto con otra cita pertinente:

“Peter Kennedy, A Guide to Econometrics, 3ª ed., The MIT Press, Cambridge, Mass., 1992, p. 82, en Gujarati p. 443

La búsqueda de los economistas por la “verdad” a través de los años ha dado origen a la versión de que los economistas son personas que andan buscando un gato negro que no existe dentro de un cuarto oscuro; con frecuencia, se acusa a los economistas de encontrar uno”

- b. Con relación a la corrección de errores, ésta es explicitada en la tabla de resultados, donde se dice que las regresiones son aplicadas con **errores robustos**. Es decir, se corrige la presencia de heteroscedasticidad.
7. Con respecto a la corrección por heteroscedasticidad y normalidad de los errores en el PSM:
- a. Lo primero a tener en consideración es conocer y saber interpretar **los errores robustos calculados por bootstrapping**; precisamente con esta técnica correctiva, los errores ya comienzan a ser robustos.

Segundo, se explicita también que la regresión es aplicada con errores robustos.

Uno de los problemas que se presentan al estimar el PSM es que al testear su significancia con base en los errores estándar, no es robusto, debido a que los errores estándar deberían incluir la varianza de la estimación del índice de propensión (Propensity Score), las imputaciones al soporte común, y también agregar información acerca de la variación asociada al pareo.

Una forma de corregir esto es mediante el Bootstrapping, como lo sugiere Lechner (2002). Éste es un popular método que permite que los errores estándar, mediante iteraciones de submuestras de los tratados, para los distintos pasos de la estimación (propensity score, matching, etc.). Con esto se garantiza que la distribución de la muestra, y así sus errores estándar, se aproximen a la media de la población. Esta metodología está incluida en el algoritmo de Ichimura y, al final de cada tabla, se dice que los errores estándares fueron calculados por Bootstrapping y, por ende, se puede garantizar errores robustos y normales.

El algoritmo utilizado los corrige de forma automática, y resulta innecesario entrar en una discusión tan técnica en un informe que no se enfoca propiamente en eso.

- b. El PSM es una técnica que se utiliza con el objeto de tener una situación contrafactual ideal para comparar el efecto de un individuo que recibe tratamiento. Por ello se utiliza una estrategia de estimación que se basa en aparear un set de variables observables entre individuos, que deberían determinar el cómo

se afecta el ingreso o la empleabilidad, pero nunca debería afectar la probabilidad de recibir tratamiento. Esto es conocido como el Supuesto de Independencia Condicional en la asignación del tratamiento. La violación de este supuesto podría generar resultados sesgados, es decir, el impacto podría ser sobre o sub valorado; por ello es de mucha importancia escoger bien las variables que influyen en la variable dependiente, y que éstas no estén correlacionadas con la variable tratamiento.

Éste es un supuesto fuerte pero válido, y que depende enteramente de la calidad y capacidad del investigador de observar las variables de interés. Un posible problema que podría aparecer en esto y que podría violar el supuesto de incondicionalidad es, por ejemplo, un vector de variables características de alta dimensionalidad, lo que podría conllevar el problema señalado en el informe. Para esto, Rosenbaun y Rubin (1983) sugieren utilizar un índice de propensión. Así, ellos demuestran que si la variable dependiente es independiente de la variable tratamiento condicional a las variables características, ellas también son independientes del tratamiento condicional a un índice de propensión balanceado (PSM). Para llegar a esto se realiza un traslape de un soporte común, que es balanceado. Por ello, se puede afirmar que el supuesto de independencia condicional se cumple.

Todo esto está muy bien detallado y explicado en el informe, en los pasos sobre cómo se estima el PSM.

Es así como la decisión de aplicar el PSM es precisamente porque esta metodología garantiza cumplir los supuestos, resolviendo los problemas de ortogonalidad en los errores.

- c. Las metodologías para solucionar problemas que enfrenta el PSM cuando el grupo de control es menor están expuestas y aplicadas, junto con las técnicas involucradas; y así lo detallan las propias explicaciones consignadas en el informe final:

“...La consideración más relevante que hay que tener según dicho trabajo, es el tamaño de la muestra. Claramente, de manera asintótica el PSM tenderá a parear al individuo tratado con su control exacto a medida que la muestra sea más grande, sin importar el algoritmo de pareo. Sin embargo, en muestras pequeñas la elección del algoritmo de pareo resulta ser muy importante (Heckman, Ichimura and Todd, 1997), lo que implica enfrentarse a un problema de costo de oportunidad entre sesgo de selección y aumento de varianza innecesario. En virtud de ello, se aplicará un pareo de Kernel, con el fin de aprovechar la característica que tienen los datos de tener muchos individuos en el grupo de control que son comparables con los tratados, es decir, que a pesar que el grupo de control pueda ser pequeño, los individuos que pertenecen a él se pueden

comparar con más de un individuo que participó en el Programa, todo esto por el diseño de la muestra que permite contar con individuos muy similares en ambos grupos...”

En ese párrafo del informe se justifica cómo se corrige el problema de la muestra menor, para lo cual se aplica el pareo por Kernel. Es importante destacar, como se hace en el informe, que con la aplicación de esta técnica, solamente quedaron fuera 5 controles y 1 estudio de entre los 1.153 individuos que conforman ambas muestras. Esto último vuelve a confirmar la enorme semejanza entre ambos grupos.

8. Finalmente, adosamos el cuadro consolidado de los resultados de impacto y sus comentarios asociados, donde lo principal es la sugerencia en relación con explorar nuevas metodologías para la medición de impactos del Programa F T.

El prisma para mirar los resultados se sintetiza en el siguiente consolidado de resultados. En este sentido, lo que verdaderamente importa es que todos los instrumentos arrojan impactos positivos. Las cifras exactas de cada uno de los indicadores hay observarlas e interpretarlas en función de los significados específicos de cada uno de ellos:

- El evaluador mide la eficiencia
- El impacto sobre la media de ingresos mide el incremento promedio real de los salarios
- El impacto sobre la inversión unitaria representa una muy importante relación a tener en consideración para medir el impacto propio desde la perspectiva de la inversión efectuada
- De los índices financieros, desde nuestro punto de vista, el período de recuperación es el más relevante, pues da cuenta del tiempo en que se recupera toda la inversión realizada en un año

<i>Instrumento</i>	PSM	MRL	Evaluador	VAN Social	Período Recuperación	TIR Anual
General	\$ 24.870	\$ 31.717				
M: \$ 315.541	7,9%	10,1%	4,4%	\$ 7.513.295.592.259	Feb-04	130%
IU: \$ 128.277	19,4%	24,7%				
Año 2002		\$ 27.120				
M: \$ 285.902		9,5%	4,6%	\$ 2.637.562.291.437	Oct-03	125%
IU: \$ 157.364		17,2%				
Año 2003		\$ 29.308				
M: \$ 304.180		9,6%	3,9%	\$ 2.462.769.142.186	Nov-04	146%
IU: \$ 140.029		20,9%				
Año 2004		\$ 38.604				
M: \$ 309.671		12,5%	5,3%	\$ 1.582.476.770.100	Jun-06	131%

IU: \$ 117.946	32,7%				
Año 2005	\$ 32.201				
M: \$ 345.759	9,3%	3,8%	\$ 703.058.771.040	Dic-06	133%
IU: \$ 109.948	29,3%				
Año 2006	\$ 42.824				
M: \$ 333.911	12,8%	5,5%	\$ 348.426.691.926	Nov-07	113%
IU: \$ 133.444	32,1%				
Año 2007	\$ 20.407				
M: \$ 314.000	6,5%	3,5%			
IU: \$ 110.930	18,4%				

- En materia de género, todos los resultados econométricos señalan que existe una diferencia consistente y sin excepción por año entre hombres y mujeres. Estas últimas ven desfavorecidos sus ingresos en relación con los hombres. La cifra que reporta la regresión global de los seis años es de \$58.043 y significativa al 1%. Los hombres ganan \$58.043, en promedio mensual, más que las mujeres, independientemente de los tipos de capacitación que efectúen o si no la hacen. Y este resultado no es aleatorio, sino que es plena y consistentemente tendencial, como lo confirman las regresiones realizadas por año.
- La experiencia internacional en materia de evaluaciones –tal como lo destaca el estudio internacional realizado en esta consultoría- señala que las magnitudes de los impactos en Programas de capacitación similares en el mundo, son bajas y sin significación estadística.
- La óptica para mirar los resultados de impacto en los Programas de capacitación debe ser tomando en cuenta las inversiones previas que se relacionan con la formación (educación formal) del capital humano base de los trabajadores y que son muy superiores. De esta manera, la capacitación actúa en la esfera de los diferenciales de salarios con relación a un determinado monto base de capital humano ya creado en la anterior etapa de la educación formal.

- En consecuencia, si se compara inversión unitaria promedio de \$128.277 con el costo unitario de la educación formal en este país (por ejemplo los doce años de una educación media completa, excluyendo los costos que implica algún tipo de educación superior), no hay ninguna relación posible en términos de órdenes de magnitud. Por lo tanto, los incrementos marginales de salarios que exhiben los resultados, hay que observarlos a la luz de la magnitud y sentido de esta inversión unitaria.
- Las metodologías actuales para medir impactos en el Programa Franquicia Tributaria son insuficientes y limitadas; solamente se centran en los diferenciales de salarios de los trabajadores capacitados.
- La sugerencia es que el SENCE estimule la creación de nuevas metodologías, dirigidas específicamente a la evaluación integral de impactos del Programa Franquicia Tributaria, a partir de las actuales metodologías empleadas.
- En esta línea sugerimos iniciar, al menos, cinco caminos nuevos, paralelos y complementarios:
 1. Generación de métodos específicos de medición de la evolución de la productividad de las empresas que utilizan la Franquicia: indagar en una modelación simple y que requiera de las empresas información que sea de fácil acceso, la cual permita observar las evoluciones de la productividad de ellas en el tiempo.
 2. Avanzar hacia la construcción de un modelo universal que represente y explique el comportamiento de los salarios de los capacitados, mediante técnicas de regresiones econométricas: Se trata de avanzar en una investigación profunda que logre especificar un modelo universal que arribe a una forma funcional definida.
 3. Diseño de modelos estocásticos con presencia de heteroscedasticidad: implica el desarrollo de modelos donde se especifican las distribuciones de probabilidad condicional junto con los momentos respectivos, es decir, la modelación de los momentos.
 4. Explorar la posibilidad de construir Modelos de Control de Impactos: este camino implica apelar a matemáticas más avanzadas y su arquitectura definitiva consistirá, lo más probable, en sistemas de ecuaciones diferenciales parciales con resolución numérica.
 5. Generar instrumentos metodológicos para la evaluación del Programa en proceso: éstos son más simples, pero de enorme interés para el SENCE, pues le permitiría monitorear el Programa en su decurso temporal.

Creemos que en torno, al menos, de estos cinco puntos, debiera girar la reflexión profunda sobre metodologías de evaluación y medición de impactos en relación con el Programa Franquicia Tributaria, en atención a su singularidad. Éstos son, en nuestra opinión, los desafíos que vienen y lo debiera estimular el Programa en materia de metodologías *ad hoc*.

Si bien metodologías como el PSM son robustas, la medición de los diferenciales de salarios de los trabajadores, como único camino para la determinación del impacto del Programa Franquicia Tributaria, es insuficiente y debe ser perfeccionada.

Más aún, las líneas de acción en lo metodológico que se sugieren, no harían sino complementar todo el instrumental actual.