

Desarrollo de una metodología para proyectar indicadores de desempeño: el caso del Plan TIC Colombia

Ricardo Bórquez C.



Este documento ha sido elaborado por Ricardo Bórquez, bajo la coordinación de Massiel Guerra y Mariana Balboni de CEPAL en el marco de la asistencia técnica solicitada por el Ministerio de Tecnologías de la Información y las Comunicaciones de la República de Colombia. Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de exclusiva responsabilidad del autor y pueden no coincidir con las de las organizaciones involucradas.

El presente documento fue preparado en el marco del Observatorio para la Sociedad de la Información en Latinoamérica y el Caribe (OSILAC), un proyecto desarrollado por la División de Desarrollo Productivo y Empresarial de la CEPAL, que recibe asistencia financiera del Centro Internacional de Investigaciones para el Desarrollo (CIID).

Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de exclusiva responsabilidad del autor y pueden no coincidir con las de la Organización.

Este documento puede ser descargado en línea en <http://www.cepal.org/>.

Índice

Resumen	5
I. Introducción	7
1. Modelo económico	7
2. Metas en base a pronósticos	9
II. Modelo Estadístico	11
1. Supuestos	11
2. Función de verosimilitud	12
3. Pronóstico de largo plazo	14
III. Definición de metas en el plan TIC Colombia	16
1. Datos	17
2. Resultados	18
3. Conclusión	19
IV. Bibliografía.....	21

Índice de figuras

Figura 1 Distribución de medidas de desempeño	9
Figura 2 Ejemplo Plan TIC Colombia	18

Resumen

En este trabajo se describe una metodología estadística diseñada para establecer metas apropiadas sobre indicadores de desempeño, esto es, sobre mediciones cuantitativas de la gestión de un individuo (o responsable). Una meta apropiada se define como una meta que por una parte reduce el riesgo moral que enfrenta un responsable y por otra, asigna al individuo más eficiente el máximo incentivo. El contexto de esta aplicación es la definición de metas que ayudarán al seguimiento y evaluación del Plan TIC Colombia, actual política nacional para el desarrollo de la Sociedad de la Información en ese país. La metodología está sustentada en un modelo económico simple donde a los responsables se les incentiva ante el evento de cumplir una meta, pero existe incertidumbre respecto del desempeño a lograr por el efecto de covariables que inciden sobre el resultado. Las metas se determinan mediante pronósticos de las variables de desempeño, bajo una condición necesaria de optimalidad sobre la eficiencia en la gestión de los responsables. La caracterización dinámica del modelo permite construir pronósticos de corto plazo como también de largo plazo, con información agrupada en un panel de variables de resultados para un conjunto de individuos. Los parámetros del modelo se estiman consistentemente y eficientemente maximizando una función de cuasi-verosimilitud y la distribución asintótica de los estimadores es normal.

In this paper, we describe a statistical methodology designed for establishing appropriate targets over performance indicators. Consider a target as appropriate if it reduces moral hazard confronted by an agent and it assigns the maximum prize to the most efficient agent. The context for introducing this application is the definition of targets for the Plan TIC Colombia, a national policy for the development of the Society of Information in that country. Our methodology is based on a simple economic model, where agents perceive an incentive in the case of achieving a target and uncertainty arises because the effect of covariables. Targets are defined using conditional forecasting of the performance variables, under a necessary condition regarding economic efficiency. The dynamic characterization of the model allows us to obtain both short term and long term forecasts, relative to a panel data of information about several performance indicators and agents. Model parameters can be estimated consistently and efficiently by maximizing a quasi-likelihood function and the asymptotic joint distribution of the estimators is normal.

I. Introducción

El Plan Nacional de Tecnologías de la Información y las Comunicaciones (en adelante Plan TIC) que ha impulsado el Ministerio de Tecnologías de la Información y las Comunicaciones de la República de Colombia ha considerado la implementación de una estrategia para el desarrollo de la Sociedad de la Información con objetivos concretos de largo plazo, a cumplir hacia el año 2019. El marco conceptual del Plan TIC Colombia ha sido elaborado siguiendo la metodología de un *balanced scorecard*. En el diseño logrado, se han propuesto dos grupos de ejes principales: ejes transversales (comunidad, gobierno en línea, investigación, desarrollo e investigación, marco normativo, regulatorio e incentivos) y ejes verticales (educación, salud, justicia y competitividad empresarial), donde para cada eje se contemplan objetivos estratégicos que serán ejecutados a través de programas y proyectos específicos¹.

Los avances de las actividades que se desarrollan en el Plan TIC son monitoreadas mediante un sistema de indicadores que se conocen periódicamente. En particular, el proceso de monitoreo se encuentra orientado hacia el cumplimiento de metas específicas definidas en cada programa o proyecto, los que permanecen a cargo de un conjunto de responsables de su ejecución.

El presente trabajo tiene por objetivo desarrollar una metodología estadística que permita establecer metas apropiadas para estos indicadores. La calificación de “apropiada” es un tanto subjetiva, por lo que se intentaría conceptualizar este término en lo que viene. Por ahora se dirá que una meta apropiada induce un desempeño óptimo (en algún sentido a definir) en la gestión de los responsables. El desarrollo de la metodología se centra fundamentalmente en el responsable de un proyecto o programa, dado que es precisamente allí donde se generan las posibilidades de éxito o fracaso de la gestión, en relación con una meta determinada y otros eventos.

1. Modelo económico

Es conveniente explicitar el problema económico que enfrenta el responsable de un programa, lo que servirá además para motivar la discusión metodológica que se desarrollará en las secciones

¹ Para ver detalles: <http://www.colombiaplantec.org.co>.

siguientes. El modelo económico en el cual está basado el desarrollo metodológico, está inspirado en el clásico problema del agente-principal, el cual muestra un conjunto de situaciones que se originan cuando un actor económico (el principal), depende de la acción o de la naturaleza o moral de otro actor (el agente), sobre el cual no tiene perfecta información, o, en otras palabras, trata las dificultades que se presentan bajo condiciones de información asimétrica, cuando el principal contrata a un agente.

Considere un simple modelo económico para el problema de decisión bajo incertidumbre que tiene este individuo (agente), al cual se le ha fijado una meta y^* respecto de una medición de desempeño para un programa que se designará por el valor y . Adicionalmente, este individuo tiene asociada una función de costo $g(y - y^*)$ que es creciente respecto del nivel de desempeño realizado por sobre la meta exigida.

Note que esta función permite caracterizar la gestión del individuo a través de su primera derivada respecto de $y - y^*$.

Sea:

$$U(y) = \begin{cases} a & \text{si } y > y^* \\ 0 & \text{si } y \leq y^* \end{cases}$$

la función de utilidad subjetiva del individuo cuyo argumento es el nivel de desempeño y , la que se supone aditiva respecto de otros beneficios. En cada período el individuo recibe un premio $a > 0$ si cumple la meta y^* o bien recibe 0 en caso contrario, con lo que debe resolver el siguiente problema de optimización²:

$$\max_y E[U(y)] = \max_y \left\{ aP(y > y^* \mid g(y - y^*) \leq C) \right\} \quad (1)$$

donde $P(y > y^* \mid g(y - y^*) \leq C)$ es la probabilidad de exceder la meta y^* dado que se satisface la restricción de máximo costo que el individuo está dispuesto a enfrentar: $g(y - y^*) \leq C$. Esta probabilidad también depende de la información que el individuo dispone acerca del desempeño pasado de los programas a su cargo, del desempeño corriente y pasado de los programas a cargo de otros individuos y de variables exógenas que puedan afectar su desempeño. Se ha omitido explicitar esta dependencia con el propósito de simplificar la notación.

La solución al problema planteado en (1) está dada por el nivel de desempeño óptimo que escogerá el individuo en consideración de los incentivos que enfrente, por lo que ésta depende explícitamente de la meta asignada y^* , de la función de costos $g(\cdot)$ y del nivel máximo de costos C . En efecto, la condición de primer orden es $-aD(y^* \mid g(y - y^*) \leq C) = 0$ donde

$D(y^* \mid g(y - y^*) \leq C)$ corresponde a la densidad condicional de y evaluada en y^* dado $g(y - y^*) \leq C$. Note que $D(y^* \mid g(y - y^*) \leq C) = \frac{D(y^*)}{P(y \leq g^{-1}(y^*, C))}$ donde $g^{-1}(\cdot)$

corresponde a la función inversa de g respecto de cualquier nivel de (y^*, C) mientras que $D(y^*)$ es la densidad marginal de y evaluada en y^* .

² La extensión hacia la optimización multiperiodo es posible, pero añade una complejidad adicional en el modelo que no es necesaria en este contexto, razón por la cual se ha descartado del análisis.

Por monotonía en $P(y \leq g^{-1}(y^*, C))$, el nivel óptimo de desempeño que escogerá un individuo que se enfrenta a un problema de decisión como el descrito en (1) está dado por $y = g^{-1}(y^*, C)$, donde la existencia de un máximo está garantizada bajo la condición de segundo orden para este problema.

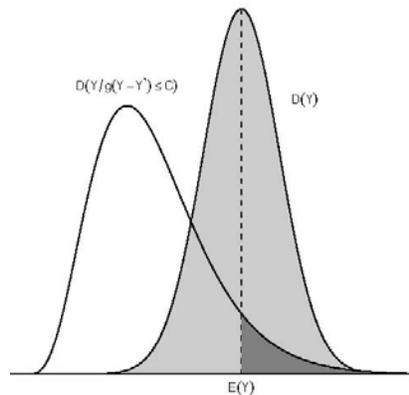
Interesa determinar las consecuencias de un cambio marginal en y^* sobre el nivel de desempeño óptimo del individuo en este modelo. Para ello, suponga que $D(y)$ puede aproximarse por una densidad simétrica (por ejemplo, la densidad normal) con media $E(y)$, esto señala que existe una baja probabilidad que el desempeño del individuo se ubique en los extremos y una alta probabilidad que sea cercano a la media. Lo que se busca es que ante un alza en el nivel de exigencia el individuo escoja óptimamente *aumentar* su nivel de desempeño. Diferenciando la expresión $-aD(y^* | g(y - y^*) \leq C) = 0$ respecto de y^* , se obtiene la siguiente relación:

$$\frac{\partial y}{\partial y^*} = D(y^* | g(y - y^*) \leq C)^{-1} \frac{\partial D(y^*)}{\partial y^*}, \text{ donde esta expresión es positiva sólo si } y^* \leq E(y).$$

2. Metas en base a pronósticos

La primera implicancia del modelo desarrollado en el punto anterior, es que debido a la restricción de costos las metas deben establecerse **por debajo** del pronóstico incondicional de desempeño $E(y)$. La razón más intuitiva es simple, si la meta es muy exigente entonces la utilidad esperada puede ser menor que el costo que conlleva un mayor esfuerzo y el resultado óptimo a nivel del individuo puede ser finalmente reducir su desempeño, lo que obviamente es contrario a lo que se busca en el contexto de la implementación del modelo. Lo que ocurre es que $P(y > E(y) | g(y - y^*) \leq C)$ que es la probabilidad que utiliza el individuo para determinar su nivel de desempeño óptimo resulta ser, para cualquier valor de $E(y)$, estrictamente menor que $P(y > E(y))$, que es la probabilidad de alcanzar la meta sin considerar los costos asociados. En otras palabras, un individuo tomará sus decisiones asignando una probabilidad más baja a los eventos de mayor desempeño de lo que ocurriría si no enfrentara una restricción de costos.

FIGURA 1
DISTRIBUCIÓN DE MEDIDAS DE DESEMPEÑO



Fuente: Elaboración propia.

Esto se ha representado en la Figura 1, donde $P(y > E(y) | g(y - y^*) \leq C)$ equivale al área destacada en color gris oscuro, mientras que $P(y > E(y))$ corresponde al área en gris claro bajo la curva $D(y)$ que está a la derecha de la línea punteada que marca la ubicación del pronóstico incondicional $E(y)$. De hecho, lo descrito previamente es un problema de asimetría de información conocido como riesgo moral, en tanto únicamente el responsable de un programa (y no su mandante) es quien conoce el nivel de esfuerzo que destinará a realizar la gestión encomendada y su decisión dependerá fundamentalmente de los incentivos económicos que enfrente.

La segunda implicancia es más evidente, si la meta es demasiado baja entonces el nivel de desempeño logrado puede no ser suficiente para cumplir con los objetivos propuestos, lo cual contradice la motivación primera que guiaba la necesidad de establecer una meta. La solución a este *trade-off* viene de un argumento estándar en teoría de la oferta, que en este caso se interpreta como una restricción de optimalidad en el esquema de incentivos, por la vía de vincular el premio al desempeño con la eficiencia en la gestión: el individuo más eficiente es quien percibe la mayor utilidad esperada sin que ello implique un detrimento en su nivel de desempeño. Note que esta es una condición necesaria que debe satisfacer un sistema de incentivos para evitar posibles distorsiones en la asignación de niveles de desempeño por parte de los responsables. Esto ocurre precisamente en el punto $y^* = E(y)$ donde la esperanza ahora se calcula desde la perspectiva del individuo más eficiente. En definitiva, una meta apropiada en nuestro contexto es una meta que:

1. Reduce el riesgo moral que enfrenta un individuo responsable, al recompensar (en términos esperados) el esfuerzo realizado por sobre los costos respectivos;
2. Asigna al individuo más eficiente el máximo incentivo.

Sólo en tal caso se puede asegurar que el responsable se encontrará debidamente motivado para llevar su desempeño en el mismo sentido que la meta propuesta para su evaluación. Es necesario señalar que en el análisis anterior se han mantenido inalterados otros parámetros del modelo y como es de esperar, un incentivo más alto ($a' > a$) o una restricción de costos más holgada ($C' > C$) pueden también ayudar a modificar la conducta del responsable en el sentido requerido. Asimismo, la incorporación de la aversión al riesgo en el análisis es directa notando que los beneficios esperados con una meta demasiado elevada pueden ser insuficientes (aún si los costos son relativamente bajos) para un individuo muy averso al riesgo.

Definido el marco teórico-económico sobre el cual deben elaborarse los pronósticos de desempeño para el establecimiento de metas apropiadas, en la sección siguiente se discute la metodología que permite su implementación práctica. El objetivo será estimar en base a la información disponible la mejor predicción acerca del nivel futuro de las variables de desempeño (resultados de programas), indicadores de desempeño individuales y también de indicadores más agregados que involucran a varios responsables.

II. Modelo Estadístico

1. Supuestos

Sea $X_T = (x_1, x_2, \dots, x_T)$ la información disponible (observada) al momento T , por ejemplo, x_1 corresponde a la información acerca del desempeño de todos los responsables y de cada uno de los programas en el primer año. En general, denote por $x_t = (y_t, z_t)'$ la información indexada al momento t , la cual es particionada en variables que se denominarán variables de *resultado* (y_t) y variables de *inducción* (z_t). Se entiende por variables de inducción aquellas que eventualmente tienen un efecto sobre las variables de resultado pero que no dependen de la gestión de los responsables de los programas. Lo que se busca es caracterizar el proceso $X_T | X_{T-1}$ (léase X_T dado X_{T-1}) con el propósito de construir un pronóstico condicional para las variables de respuesta de la forma $E [Y_{T+1} | X_T; \psi] = E [Y_{T+1} | x_1, x_2, \dots, x_T; \psi]$, donde ψ , es el parámetro de interés del modelo y Y_T agrupa a las variables de respuesta hasta el momento T (respectivamente Z_T agrupa a las variables de inducción). La justificación de los supuestos necesarios para construir dicho pronóstico se discute a continuación.

Cada responsable tiene a su cargo un conjunto de programas cuyo desempeño se mide en variables que se han agrupado en X_T , algunas de las cuales serán rotuladas como variables de resultado mientras que otras como variables de inducción. Las variables de resultado tienen la característica que en cada momento t su evolución en el tiempo depende de su propia historia Y_{t-1} y también de la trayectoria pasada de las variables de inducción Z_{t-1} . La dependencia temporal está justificada en las acciones que se aplicarán periódicamente para adecuar el resultado corriente a las metas propuestas, lo que se traducirá en última instancia en algún esquema implícito o explícito de incentivos sobre la gestión de los responsables. Por esa misma razón, es esperable que existan diferencias en el modelo condicional a nivel de los responsables y en particular que el desempeño de variables que dependen de un mismo individuo se encuentren estadísticamente relacionadas entre ellas. Esto refleja la necesidad de distinguir entre parámetros estructurales del modelo (agrupados en λ) y parámetros incidentales (cuya dimensión varía con el número de individuos).

La formalización matemática de estos conceptos está dada por: para todo t se tiene que $Y_{t-1} \parallel z_t | Z_{t-1}, \lambda, \theta$ (el símbolo \parallel denota independencia condicional) donde $Y_{t-1} = (y_1, \dots, y_{t-1})'$, $Z_{t-1} = (z_1, \dots, z_{t-1})'$ y $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ es un parámetro incidental; ii) $(y_t | z_t; \lambda_1, \theta)$ y $(z_t; \lambda_2)$ operan un corte secuencial respecto de $\lambda = (\lambda_1, \lambda_2)$ y iii) $\psi \subset \lambda_1$. El primer supuesto señala que las trayectorias

observadas en el pasado para las variables de resultados no aportan información relevante en relación al pronóstico (condicional) de las variables de inducción. El segundo supuesto dice que la inferencia respecto a los parámetros del modelo condicional no depende del proceso particular que describe a las variables de inducción. El tercer supuesto establece que el parámetro de interés está presente únicamente en la densidad condicional $y_t|z_t$, por lo que z_t puede considerarse una variable exógena. Más aún, con el supuesto i) la variable z_t es fuertemente exógena (ver Engle *et al.* (1983)). Bajo estas condiciones la densidad de $Y_T|X_{T-1}$ se puede escribir como sigue:

$$D(Y_T | X_{T-1}; \lambda, \theta) \propto \prod_{t=1}^T D(y_t | z_t, X_{T-1}; \lambda_1, \theta)$$
 y más importante, la inferencia respecto de ψ se puede conducir obviando la densidad de Z_T sin pérdida de información relevante.

2. Función de verosimilitud

Para operacionalizar esta propuesta es necesario definir primero la forma funcional de $D(Y_T | X_{T-1}; \lambda, \theta)$, por lo se debe precisar el rol que juega el efecto aleatorio en este modelo. Más específicamente, se tiene para todo t que $y_{ijt} \parallel y_{ij't} | z_{ijt}, X_{t-1}, \lambda, \theta_i$ dado $j \neq j'$ con lo que

$$D(Y_t | X_{t-1}; \lambda, \theta) \propto \prod_{i=1}^n \prod_{j=1}^{J_i} D(y_{ijt} | z_{ijt}, X_{t-1}; \lambda_1, \theta_i)$$
 donde $i = 1, 2, \dots, n$ indexa a los distintos

responsables y $j = 1, 2, \dots, J_i$ indexa a los programas a cargo de cada responsable. Luego, la relación contemporánea en el resultado de dos programas asociados a un mismo individuo controlando por la información disponible hasta $t-1$ y el efecto de las variables de inducción (exógenas), depende de cada responsable sólo a través del efecto aleatorio y en consecuencia, es posible introducir por esta vía la restricción de optimalidad discutida en el marco teórico-económico en la forma que se describe más adelante.

Siguiendo a Papke y Wooldridge (1996) para la estimación de los parámetros se especifica una función de cuasi-verosimilitud basada en un modelo Bernoulli con parámetro $p_{ijt} = D(y_{ijt} | z_{ijt}, X_{t-1}; \lambda_1, \theta_i)$ para cada variable (o resultado de un programa). Este método es eficiente respecto de una amplia clase de modelos y produce estimadores que asintóticamente se distribuyen normalmente³. Si bien en este contexto se restringe la modelación a variables que toman valores únicamente en el intervalo (0,1), también se simplifica significativamente la especificación estadística⁴. Se modela p_{ijt} como una función de distribución logística evaluada en covariables $A'_{(ij)} \text{vec}(X_{t-1})$ donde $A_{(ij)} \subset \psi$ define la estructura de parámetros que es relevante para explicar a la j -ésima variable de resultado asociada al i -ésimo individuo⁵. La ventaja de esta función es que no impone restricciones sobre el espacio paramétrico de ψ .

³ Si el propósito es realizar test de hipótesis sobre los parámetros los autores proponen hacer inferencia robusta asintóticamente, para lo cual se requiere de una corrección en la matriz de varianza-covarianza de los estimadores.

⁴ La elección de una transformación apropiada de los datos depende de las características estadísticas de la serie original y consecuentemente puede afectar la consistencia de los estimadores de los parámetros del modelo, lo que se discute más adelante. El caso más simple corresponde a un proceso estacionario al cual se le aplica una simple transformación de escala, lo que no altera las propiedades estadísticas del pronóstico.

⁵ La notación aquí puede ser algo confusa por lo que es conveniente hacer una aclaración. Las dimensión de la matriz $A_{(ij)}$ no varía con los subíndices sino que su composición es específica a cada programa, en el sentido de que identifica cuáles son los parámetros estructurales que ayudarán a explicar el desempeño futuro de ese programa, con elementos iguales a 0 en todas las otras posiciones. En otras palabras, esta matriz resulta de imponer restricciones de no-dependencia sobre los parámetros estructurales del modelo y que son aplicables a cada programa en particular.

El número de parámetros a estimar es un problema importante a considerar en el contexto dinámico del modelo, debido a la escasa cantidad de datos disponibles para cada variable en la dimensión temporal del panel. No obstante, la especificación más general está basada en la siguiente función de cuasi-verosimilitud:

$$\ell(\psi | X_T) = E \left[\prod_{t=1}^T \prod_{i=1}^n \prod_{j=1}^{J_i} p_{ijt}^{y_{ijt}} (1 - p_{ijt}^{y_{ijt}}) \right] \quad (2)$$

donde $p_{ijt} = \frac{1}{1 + \exp(-A'_{(ij)} \text{vec}(X_{t-1}))}$ y la esperanza se calcula integrando respecto de la densidad (condicional en Z_T) del efecto aleatorio (ver Wooldridge (2002), p. 412).

La condición de optimalidad económica impone que el modelo se deba conceptualizar desde la perspectiva del individuo más eficiente, lo que se formula utilizando la densidad de $\max_{\theta_i \in \Theta} \theta_i | Z_T$ (el máximo de $\theta_i | Z_T$) al momento de integrar la expresión en (2). La densidad asociada a un estadístico de orden es asimétrica por lo que es claro que la restricción de eficiencia opera por la vía de asignar probabilidades más altas a niveles de desempeño más altos, que es precisamente lo que se esperaría de un individuo más eficiente.

La identificabilidad de los parámetros está sujeta a restricciones de rango usuales sobre la secuencia $\{A_{(ij)}; i = 1; \dots, n, j = 1, \dots, J_i\}$ y su estimación se realiza de manera estándar resolviendo las ecuaciones de verosimilitud $\frac{\partial \ell(\psi | X_T)}{\partial \psi} = 0$ con $\hat{\psi} = \arg \max_{\psi \in \Psi} \ell(\psi | X_T)$ mediante un

procedimiento que requiere de aproximación numérica y una subrutina para evaluar en cada paso la integral respecto del efecto aleatorio. Falta todavía señalar la forma en que se obtienen los pronósticos de cada variable de resultado, pero basta notar que una vez que han sido estimados todos los parámetros del modelo, el pronóstico a un período se puede calcular simplemente como sigue: $E[y_{ijt+1} | X_T; \hat{\psi}] = \hat{p}_{ijt+1}$ donde la última expresión sólo depende de variables observadas al momento T . Este valor constituye la mejor aproximación -dada la información disponible- al valor que tomarán las distintas variables de respuesta el próximo período⁶. La formulación de un pronóstico a más largo plazo se discute en detalle en la siguiente sección.

El modelo desarrollado hasta ahora se sustenta en una formulación estadística de los resultados de un conjunto de programas administrados por distintos individuos. Pero si el interés del pronóstico se centra más bien en los responsables, por ejemplo, a través de un indicador de desempeño individual, entonces la especificación anterior puede resultar ineficiente al construir separadamente para cada programa los pronósticos respectivos. Una especificación diseñada para pronosticar indicadores de desempeño a nivel del individuo es la siguiente. Sea $y_i = (y_{i1}, y_{i2}, \dots, y_{iJ_i})'$ el vector de resultados asociado al individuo i y defina $\tilde{y}_i = \omega'_i y_i$ como el indicador de desempeño para este individuo (esto es simplemente una suma ponderada de los resultados de cada uno de los programas a su cargo), donde $\omega_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{iJ_i})'$ es un vector *conocido* de ponderadores para cada resultado de programa tal que $\omega'_i \mathbf{1}_{J_i} = 1$ (donde $\mathbf{1}_{J_i}$ denota al vector unitario de dimensión J_i). El objetivo es construir un pronóstico condicional para este indicador, el cual está basado en una descripción de la probabilidad conjunta de los resultados de los programas respectivos y que

⁶ Un pronóstico como el propuesto que se basa explícitamente en la información disponible para su cálculo se denomina formalmente un pronóstico condicional. Existe también la alternativa incondicional que utiliza información distribucional del set de información, esto es, se basa en valores esperados y no en observaciones. No obstante, bajo el supuesto de exogeneidad para Z_T la información que se deriva de uno u otro contexto es equivalente.

también se basa en la función de distribución logística, con lo que la función de cuasi-verosimilitud resultante en este caso es:

$$\ell(\psi | X_T) = E \left[\prod_{t=1}^T \prod_{i=1}^n p_{ijt}^{\tilde{y}_{ijt}} (1 - p_{ijt}^{\tilde{y}_{ijt}}) \right] \quad (3)$$

donde $p_{it} = \frac{1}{1 + \exp(\theta_i - (B'_{(i)} \text{vec}(X_{t-1})))}$ y la esperanza se calcula respecto de la densidad de

$\max_{\theta_i \in \Theta} \theta_i | Z_T$ con el objeto de incorporar la restricción de optimalidad sobre la eficiencia de los responsables. La matriz $B_{(i)}$ cumple el mismo rol que la matriz $A_{(ij)}$ en la verosimilitud (2) pero en general contiene menos parámetros⁷.

Esta especificación utiliza más información acerca de la relación conjunta de los programas asociados a un mismo responsable de la que se obtendría al ponderar individualmente los pronósticos de las variables conseguidos a partir de la función de verosimilitud en (2), razón por la que existe una ganancia en eficiencia en la estimación de los parámetros del modelo. La forma de calcular el pronóstico a un período para este indicador viene dada por: $E[\tilde{y}_{iT+1} | X_T; \hat{\psi}] = \hat{p}_{iT+1}$.

Considere por último la construcción de un indicador aún más agregado, que involucra a un conjunto I de individuos y que se calcula de la siguiente manera: $\tilde{y} = \sum_{i \in I} b_i \tilde{y}_i$ tal que $\sum_{i \in I} b_i = 1$, con

b_1, b_2, \dots conocidos. El nuevo indicador \tilde{y} no demanda más información que la recabada desde la verosimilitud en (3) con lo que se puede obtener eficientemente un pronóstico condicional para este indicador simplemente redefiniendo los ponderadores asociados a los distintos programas. La razón es que la relación temporal en las variables especificada en (3) se sustenta en la acción de los responsables y más específicamente en su racionalidad económica. Los efectos de esto se traducen en información que se ha descrito a través de la estructura de dependencia conjunta de los programas y que ya incorpora posibles interacciones entre los resultados de programas a cargo de distintos individuos. De esta forma, construir un pronóstico condicional para la medición de desempeño de un individuo o de un grupo de individuos es, en este caso, totalmente equivalente desde el punto de vista de la eficiencia en el uso de la información. En este último caso, el pronóstico a un período

correspondiente al indicador agregado se obtiene a partir de: $E \left[\sum_{i \in I} b_i \tilde{y}_{iT+1} | X_T; \hat{\psi} \right] = \sum_{i \in I} b_i \hat{p}_{iT+1}$.

3. Pronóstico de largo plazo

En esta sección se propone cómo utilizar la información proveniente del modelo estadístico para establecer metas de largo plazo para indicadores de desempeño a través de pronósticos de largo plazo. Por la discusión previa, basta con estudiar el problema a través de indicadores a nivel del individuo ya que la extensión hacia indicadores más agregados es directa, con lo que de aquí en adelante se refiere por “información” a una especificación de verosimilitud como la descrita en (3).

Se define el pronóstico (condicional) de largo plazo para el indicador de desempeño individual \tilde{y}_i como sigue: $E[E[\tilde{y}_{iT+k+1} | X_T; \hat{\psi}]] = E[\tilde{y}_{iT+k+1} | Y_T; \hat{\psi}]$ con $k = 1, 2, \dots$ donde la

⁷ Otra alternativa para definir la función p_{it} es la función de distribución logística multivariada propuesta en Malik y Abraham (1973).

primera esperanza se calcula respecto de la densidad de las variables de inducción Z_T . El paso a seguir consiste en construir un estimador consistente para la secuencia $\{E[\tilde{y}_{iT+k+1} | Y_T; \hat{\psi}]; k = 1, 2, \dots\}$.

Bajo los supuestos especificados previamente esto es posible ya que la inferencia sobre los parámetros del modelo condicional no depende de la forma específica de la densidad de Z_T , de tal forma que cualquier trayectoria razonable para esta variable puede ser utilizada para calcular el pronóstico condicional de largo plazo de las variables de respuesta (donde el término “razonable” aquí se refiere a que dicha trayectoria se ha obtenido haciendo un uso apropiado de la información disponible a través de un método estadístico).

La metodología que se describe a continuación se puede implementar sin dificultad una vez que ya han sido estimados todos los parámetros de interés del modelo, los cuales se agrupan en $\hat{\psi}$. La idea es simple, en una primera etapa se obtiene un total de L trayectorias independientes para la variable exógena correspondientes a los siguientes k periodos sobre los cuales se construirá el pronóstico y en una segunda etapa, se evalúa el pronóstico del indicador de desempeño condicional a la l -ésima trayectoria $Z_{T+k}^{(l)}$ con $l = 1, 2, \dots, L$ y a las realizaciones pasadas del indicador de desempeño, información que se agrupa en $X_{T+k}^{(l)}$. Formalmente, esto equivale a encontrar una solución por sustitución recursiva a la ecuación en diferencias definida por $E[\tilde{y}_{iT+k+1} | X_{T+k}^{(l)}; \hat{\psi}] = \hat{p}_{iT+k+1}^{(l)}$ con $k = 1, 2, \dots$ (Pindyck y Rubinfeld (1998) se refieren a este método como “modelo de simulación”). El valor del pronóstico de largo plazo corresponde al estimador de Monte Carlo de esta esperanza que está definido por $\frac{1}{L} \sum_{l=1}^L E[\tilde{y}_{iT+k+1} | X_{T+k}^{(l)}; \hat{\psi}] = \frac{1}{L} \sum_{l=1}^L \hat{p}_{iT+k+1}^{(l)}$ con $k = 1, 2, \dots$ el cual constituye una extensión simple al método descrito en Pindyck y Rubinfeld.

III. Definición de metas en el plan TIC Colombia

Tal como se describe en la introducción de este documento, el seguimiento y evaluación del Plan TIC Colombia se realiza mediante un sistema de indicadores de desempeño sobre programas a cargo de distintos responsables, para los cuales se establecen metas concretas que buscan garantizar una adecuada implementación de la estrategia en el largo plazo. En esta sección, se implementan los procedimientos desarrollados anteriormente con los siguientes objetivos específicos: i) discutir mediante un ejemplo la aplicación práctica de la metodología para establecer metas sobre indicadores de desempeño para un conjunto de programas del Plan TIC a nivel de responsables y ii) obtener una proyección de largo plazo para estos indicadores.

1. Datos

La información disponible para la presente aplicación corresponde a resultados semestrales (entre los años 2008 y 2009) para un conjunto de veinticuatro indicadores de desempeño relativos a once programas que componen el Plan TIC Colombia. Estos programas abarcan los seis ejes estratégicos contemplados en el respectivo *balanced scorecard* (Educación, Salud, Justicia, Competitividad Empresarial, Comunidad y Gobierno en Línea). En el ejercicio no se incluye información de covariables exógenas y se asume que cada programa depende de un único responsable (en total once). Algunos programas de la muestra tienen asociados varios indicadores de desempeño mientras que otros programas sólo cuentan con una medición.

Lo anterior conforma un panel de datos que, sin embargo, no es directamente utilizable para el propósito de la estimación de los parámetros del modelo. Lo que ocurre es que en un contexto general donde las distintas variables de resultado pueden tomar valores en diferentes rangos, desarrollar un procedimiento de estimación que garantice la obtención de estimadores consistentes resulta extremadamente difícil, sino imposible para ciertas aplicaciones. De tal forma, que el problema de inferencia subyacente debe simplificarse y para ello se ha optado por dos requerimientos: primero, los datos asociados a las variables de resultado deben ser transformados con el objeto de que todos los valores se encuentren acotados al intervalo (0,1) de manera consistente con la definición del pronóstico condicional y; segundo, la modelación estadística se realiza sobre las diferencias temporales en las variables de respuesta transformadas.

La solución al primer requerimiento se puede lograr de varias maneras, pero buscamos en particular una transformación inyectiva tal que una vez obtenido el pronóstico para esta nueva variable permita —a partir de la función inversa de la transformación— reconstruir un pronóstico en la escala

original de los datos. La transformación que se propone es $h(x) = \frac{1}{1 + \exp(-x/\sigma)}$ que corresponde

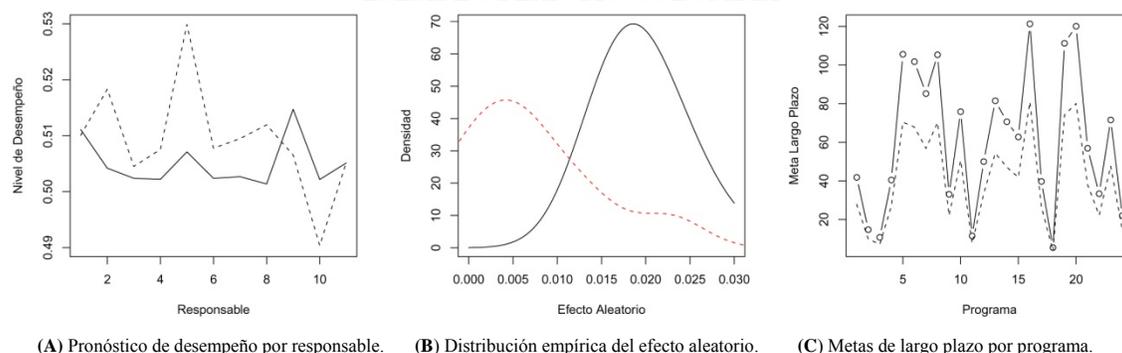
a la función de distribución logística con parámetro de escala σ . Para estimar el parámetro de escala se utiliza el estimador de varianza muestral sobre los datos originales, aunque también es posible estimarlo conjuntamente con el resto de los parámetros del modelo.

Respecto al segundo requerimiento, en la perspectiva temporal, cada variable de respuesta se cataloga como una serie integrada de primer orden debido a la forma en que se construyen los indicadores en este contexto⁸, lo que inhabilita la teoría asintótica que subyace al procedimiento propuesto por Papke y Wooldridge. Esto se soluciona si el método de estimación y pronóstico se aplica al proceso en diferencias que describe los cambios temporales en la variable de respuesta transformada⁹.

2. Resultados

El modelo representado por la función de verosimilitud en (3) ha sido ajustado a los datos disponibles, donde la respuesta asociada a cada responsable corresponde al promedio simple de las respectivas variables de respuesta (transformadas por la función $h(\cdot)$ como se detalla en la sección anterior) de los programas a su cargo. La dinámica temporal se ha descrito usando la siguiente especificación: $B_{(i)}\text{vec}(X_{t-1}) = \beta_i h(y_{t-1})$, donde β_i con $i = 1, 2, \dots, 11$ son parámetros a estimar mientras que la distribución del efecto aleatorio para cada individuo se asume normal (con lo que la distribución del efecto aleatorio del individuo eficiente es la distribución del estadístico de orden máximo de un conjunto de variables con distribución normal). El análisis de esta información nos permitirá aclarar conceptualmente la diferencia entre un pronóstico condicional de desempeño y una meta basada en dicho pronóstico. En base a los parámetros estimados se ha obtenido el pronóstico condicional de desempeño para cada responsable asumiendo que la gestión es independiente entre individuos, lo que se muestra en la Figura 2A.

FIGURA 2
EJEMPLO PLAN TIC COLOMBIA



Fuente: Elaboración propia.

⁸ Una serie integrada de primer orden se obtiene acumulando (sumando) realizaciones pasadas de una variable aleatoria cuya distribución de probabilidades no depende del índice tiempo.

⁹ Debido a que la dimensión temporal del panel es reducida, no se aplica en este caso la implementación de una especificación de corrección de errores (co-integración).

Con línea continua se ha representado el nivel de desempeño (estandarizado por su desviación estándar) que fuera observado para cada responsable durante el último semestre de 2009 y con línea discontinua, el respectivo pronóstico condicional para el semestre siguiente. Ese pronóstico toma en consideración el impacto que tiene la gestión de los responsables sobre los resultados, la cual depende del efecto de covariables. En el ejemplo, se pronosticó el mayor nivel de avance en la medición de desempeño para el siguiente semestre en el individuo N°5 quien tiene 3 programas a su cargo, mientras que el menor desempeño lo alcanzaría el individuo N°10 quien administra 2 programas. Aunque la interpretación de los resultados es limitada por la escasa disponibilidad de datos, no obstante es posible señalar que dicho pronóstico corresponde a la mejor aproximación a lo que sería el desempeño futuro de cada responsable si **no** se interviene sobre la gestión de los mismos. El análisis anterior supone que la medida de dispersión (desviación estándar) es adecuada para “homologar” los grados de avance en cada programa. Desde luego que otras medidas pueden usarse alternativamente para ese propósito dependiendo del contexto.

Por su parte, el establecimiento de metas sobre indicadores de desempeño y la retroalimentación que conlleva verificar el cumplimiento de las mismas tiene por objetivo precisamente influir sobre los resultados futuros. El problema está en establecer una meta adecuada dado que ya se conoce la dinámica temporal acerca del desempeño futuro de los distintos individuos. Como se vio anteriormente, una meta adecuada se percibe como alcanzable por el responsable disminuyendo la posibilidad de riesgo moral, a la vez que asigna el mayor beneficio al individuo más eficiente. En la Figura 2B se muestran los resultados del procedimiento de estimación en lo relativo a la gestión de los individuos (efecto aleatorio). Con línea discontinua se ha representado la densidad empírica del efecto aleatorio para los distintos responsables, mientras que con línea continua se ha dibujado la densidad teórica del individuo eficiente que es consistente con los datos. Los valores en el eje de abscisas están expresados en unidades de cambio en el nivel de desempeño de la variable estandarizada (este es el equivalente empírico de la Figura 1).

De acuerdo con el marco económico desarrollado en la sección I, la meta se debe establecer de manera asociada al pronóstico incondicional de desempeño respecto del individuo eficiente. De esta forma, la gestión efectiva debe ser recompensada en función del cumplimiento de una meta que incorpore este grado de eficiencia. En tal caso, el modelo señala que es muy probable que el individuo N°5 alcance esa meta y del mismo modo nos dice que de no intervenir en la gestión, es muy probable que el individuo N°10 no logre tal objetivo. Tal información es de evidente interés para dirigir esfuerzos que busquen el cumplimiento de las metas en todos los programas.

El pronóstico de largo plazo es sencillo de calcular en este ejemplo pues bajo los supuestos establecidos depende únicamente del parámetro de gestión asociado al individuo eficiente. Para obtener la meta en la escala original de cada variable es necesario invertir la transformación de la variable de respuesta. Este cálculo depende del parámetro de escala utilizado para estandarizar la variable. La Figura 2C muestra la meta de largo plazo para cada uno de los 24 programas evaluados, expresada como proporción del desempeño observado para el último semestre de 2009. A modo de sensibilizar este resultado, con línea discontinua se ha representado la misma meta pero usando una varianza 50% menor.

3. Conclusión

En diversas aplicaciones es relevante establecer metas sobre indicadores de gestión para el monitoreo de estrategias u objetivos. La dificultad que subyace en este contexto dice relación con el problema de agencia que se origina en la delegación de actividades en un conjunto de individuos responsables, donde cada cual decide autónomamente el nivel de esfuerzo que destinará a una actividad en particular. Esa autonomía coloca en riesgo eventual la consecución de los objetivos globales si el nivel de esfuerzo resuelto por los agentes no resulta suficiente.

El rol que juegan los incentivos es fundamental para resolver este problema, pero la implementación práctica de esta solución es usualmente complicada, especialmente si los resultados dependen de diversos factores, algunos de los cuales no están bajo el control de los agentes. En tal caso, el uso de métodos estadísticos para ayudar a establecer metas adecuadas resulta una herramienta poderosa y bien situada en dicho contexto.

El método desarrollado en este trabajo permite proveer de información importante en el monitoreo de indicadores de desempeño. Se han estimado metas asociadas a un grupo de variables de desempeño contempladas en el Plan TIC Colombia, siendo destacable la identificación de responsables cuya gestión requiere de alguna forma una intervención tal que permita el cumplimiento de metas de corto y largo plazo. La precisión de las estimaciones de metas y pronósticos dependerá en gran medida de la información disponible acerca de la relación estadística entre las distintas variables, así como de la existencia de una base de datos desagregada a nivel de responsables.

IV. Bibliografía

- Engle, R., Hendry, D. y Richard, J.F. (1983). "Exogeneity". *Econometrica* 51, pp. 277-304.
- Papke, L. y Wooldridge, J. (1996). "Econometric Methods for Fractional Response Variables With Application to 401 (K) Plan Participation Rates". *Journal of Applied Econometrics* 11, pp. 619- 632.
- Malik, H. y Abraham, B. (1973). "Multivariate Logistic Distributions". *The Annals of Statistics* 1, pp. 588-590.
- Wooldridge, J. (2002). "Econometric Analysis of Cross Section Data and Panel Data". The MIT Press, Cambridge.
- Pindyck, R. y Rubinfeld, D. (1998). "Econometric Models and Economic Forecasts". McGraw-Hill, New York.