

Software Interactivo: Observatorio Social – Económico como sistema de información y su metodología de construcción

Gonzalo Villa Cox^a, Alicia C. Guerrero^b, Ramón Villa Cox^c

^a Facultad de Economía y Negocios, Escuela Superior Politécnica del Litoral, Km. 30.5 Vía Perimetral, Guayaquil, Ecuador
gonzalo.villa@barcelonagse.eu

^b Centro de Desarrollo Social Aplicado, CEDESA, de la Facultad de Economía y Negocios, Escuela Superior Politécnica del Litoral, Km. 30.5 Vía Perimetral, Guayaquil, Ecuador
aguerrer@espol.edu.ec

^c Centro de Investigaciones Económicas de la Facultad de Economía y Negocios, Escuela Superior Politécnica del Litoral, Km. 30.5 Vía Perimetral, Guayaquil, Ecuador
ramon.villa.cox@gmail.com

Resumen

Con la finalidad de poder realizar investigaciones socioeconómicas que contribuyan a un análisis de la realidad nacional actual con miras a recomendaciones de políticas públicas, fortaleciendo así los vínculos con la comunidad a través de la academia, la Facultad de Economía y Negocios de la ESPOL, a través de sus centros, CEDESA y CIEC, decidió crear e implementar el software denominado Observatorio Económico Social. Este constituye una plataforma informática que calcula diversos indicadores sociales basados principalmente en la variable empleo, los cuales a su vez pueden ser desagregados por género, educación, etnia, zona urbana, rural y edad. Si bien es cierto, en Ecuador existen otros sistemas de información que cumplen la misma función brindando incluso una base más amplia de indicadores, el Observatorio Económico Social contribuye a realizar investigaciones más robustas, al calcular estimadores estadísticos como la varianza, pudiendo realizar así por ejemplo pruebas de hipótesis. El presente documento tiene como objetivo detallar el proceso de construcción del Observatorio Social Económico como sistema de información y consulta, aplicando para ello tanto herramientas ingenieriles como econométricas, buscando así realizar un aporte a la comunidad.

1. Introducción

A partir del año 2006, en Ecuador se han iniciado varios cambios estructurales principalmente en materia de política social y económica, para definitivamente en el 2008 implementar como modelo económico el “Modelo Social Solidario”¹. Se han emprendido así diversos proyectos y reformas destinadas a mejorar el sector de la salud, educación, vivienda y del empleo por parte del Gobierno Nacional, lo que se ha

¹ Constitución de la República del Ecuador aprobada en el 2008, artículo 283.

traducido en un aumento del 50% del presupuesto destinado al sector social devengado per cápita. Se tiene así que para el 2005 dicho valor per cápita era de \$299,6, siendo para el 2011 de \$439,8 según el Ministerio de Economía y Finanzas.

En base a esto la Facultad de Economía y Negocios, a través del Centro de Desarrollo Social Aplicado (CEDESA) y el Centro de Investigaciones Económicas (CIEC), decide construir el “Observatorio Social Económico”, como herramienta de consulta y desarrollo de investigación que brinda a la comunidad en general una serie de indicadores y estadísticas que permitan caracterizar la situación de los ecuatorianos principalmente en lo referente a educación y acceso al mercado laboral.

Se busca así brindar un software que permite calcular una serie de indicadores socio - económicos para evaluar el progreso que se ha realizado en los diferentes aspectos previamente mencionados. Cada uno de los indicadores construidos en base a las principales variables de estos sectores, será también estimado en distintos niveles de desagregación según la necesidad de información del usuario de la plataforma.

Para ello se diseñó un modelo econométrico que permita estimar de manera versátil cada indicador a los niveles de desagregación requeridos por consulta, pero que a su vez sea robusta a cualquier estructura de heterocedasticidad presente por las desagregaciones realizadas.

Los criterios de desagregación permitidos en la plataforma son: género, grupos de edad, etnia, área, región natural, nivel de instrucción y las 10 ciudades más importantes (considerando como categoría aparte el resto del país). Debido a que si se realizaran todas o varias de estas desagregaciones de forma simultánea la submuestra resultante sería muy pequeña para una estimación representativa de la variable, se decidió que la plataforma permita una elección de 2 de los criterios mencionados al mismo tiempo.

El principal aporte de la plataforma construida a lo ya calculado por el INEC y otros organismos como Sistema Integrado de Indicadores Sociales del Ecuador (SIISE), está en la estimación de indicadores robustos que permitan realizar inferencia a lo largo del tiempo y entre diferentes grupos de la población. Es por esto que además de la estimación de cada indicador se incluye su varianza ya que es una herramienta esencial para que un investigador realice inferencia estadística.

El insumo base para la elaboración de la plataforma de consulta de indicadores es la Encuesta Nacional de Empleo y Desempleo, ENEMDU. Esta es recolectada con periodicidad trimestral y con representatividad nacional principalmente en diciembre (aunque en algunos años también se ha realizado con representatividad nacional en otros trimestres como por ejemplo junio). En los otros trimestres la encuesta solo es realizada con representatividad urbana. Dado que la ENEMDU ha sufrido varios cambios metodológicos se decidió, con el equipo de investigadores, realizar la construcción de indicadores tomando el 2003 como año de inicio, ya que antes se

trabajaba con una diferente caracterización de cuando se considera a un centro poblado como rural o urbano.

A pesar de que se decidió construir los indicadores a partir del 2003, el cambio más importante de metodología para toma de la muestra y definición de variables a analizar a partir de la ENEMDU aplicado por el INEC se dio a partir del 2007.

Considerando que uno de los objetivos del presente proyecto es el diseño de una plataforma de consulta de indicadores que permita analizar la evolución de los mismos a lo largo del tiempo, fue necesario realizar la homogenización de las diferentes bases de datos en todo el periodo analizado, lo que requirió una inversión significativa en recursos tanto técnicos, académicos y de tiempo. Esto implicó la verificación de que los cambios realizados en la metodología de cálculo de las variables no creen una diferencia estadística significativa en los indicadores de ahí construidos y, en caso de que esta diferencia no exista, re-categorizar las variables para que sean comparables en el tiempo. Cuando se verificó que el cambio en la metodología utilizada por el INEC es tan grande que se hace imposible la comparación de una variable a lo largo del tiempo, se realizó un quiebre en la serie. Se deja claro así para el usuario de la plataforma que el análisis debe limitarse a los periodos donde esta tiene fundamentación estadística.

Finalmente cabe recalcar que este software tiene como objetivo facilitar la extensión de las series de tiempo a medida que nuevas fuentes de información secundarias sean publicadas por los organismos encargados de su recolección. Para ello se diseñó un protocolo para la homogenización de las nuevas bases de datos, para de esta forma poder ser añadidas a la plataforma y de esta forma extender el periodo de su análisis.

El presente documento se encuentra dividido en IV secciones. En la sección II se presenta tanto el resumen del tratamiento estadístico y la minería de datos realizada sobre las ENEMDU como también el modelo econométrico utilizado como base para la estimación robusta de los indicadores propuestos. En la sección III se presenta los resultados y productos hallados a partir del software a partir de una muestra de indicadores calculados y se termina el presente documento con las conclusiones respectivas. Una explicación de los diferentes indicadores calculados hasta el momento se presenta en el apéndice.

2. Metodología Aplicada

2.1 Descripción y homogenización de la base de datos.

Como fue indicado anteriormente, la Encuesta Urbana Nacional de Empleo y Desempleo (ENEMDU) ha sufrido varios cambios metodológicos en la forma de medir las diferentes variables empleo y desempleo. Y, considerando que uno de los objetivos del presente proyecto es el diseño de una plataforma de consulta de

indicadores que permita analizar la evolución de los mismos a lo largo del tiempo, fue necesario homogenizar las diferentes bases de datos en todo el periodo analizado.

Se tiene así que se han presentado dos cambios tan importantes en la metodología utilizada por el INEC en la última década que, si no se los toma en consideración, cualquier estimación que se haga para todo el periodo sufriría de un serio sesgo. El primero de estos ocurrió en diciembre del 2003 ya que antes de esta fecha solo la población urbana era analizada. A partir de esta fecha la definición de lo que constituye un centro urbano fue cambiada también. Es por esto que para la elaboración del observatorio solo se trabajara desde diciembre del 2003.

El segundo grupo de cambios metodológicos introducidos por el INEC ocurrieron en junio del 2007, representados por las modificaciones en las definiciones y clasificaciones utilizadas para el mercado laboral.

Los principales cambios en esta metodología se detallan a continuación:

1. **Definición de área urbana y rural:** En lo que se refiere a la definición de área residencial, para el periodo 2000 – 2002, se define como zona urbana como "**ciudad**" a aquellos asentamientos de 5.000 o más habitantes, y como zona "rural" o "**campo**" a las zonas que se encuentren en las periferias de las ciudades que tengan menos de 5.000 habitantes. A partir del 2003, el INEC cambió la definición de ciudad, estableciendo que ésta está conformada por 2000 y más habitantes. En esta definición, como en la anterior, las periferias de las ciudades son consideradas rurales.
2. **Periodicidad de la encuesta:** A partir de septiembre del 2003, se realiza la Encuesta Urbana de empleo y desempleo con periodicidad trimestral, sin embargo, para los 3 primeros trimestres, la representatividad es solo urbana (EUED), siendo la representatividad nacional para el último trimestre. Sin embargo, cabe resaltar, que para las encuestas realizadas en el mes de junio (III trimestre) de los años 2008 y 2010-2012 la representatividad también fue nacional. Parece que este patrón se va a mantener en el futuro.
3. **PEA según mercado laboral:** El mercado es segmentado a partir del 2007 en 3 sectores: Sector Formal, Sector Informal y Servicio Domestico. Antes de esto era segmentado en: Sector Moderno, sector Informal, Actividades Agrícolas y Servicio Doméstico. Esto fue realizado reconociendo que un trabajador del sector de las Actividades Agrícolas puede encontrarse tanto en el sector formal o el informal. El sector informal es medido a través de la legalidad del mismo, esto implica la existencia de RUC y registros contables completos.

4. **El periodo de referencia para la búsqueda de empleo** es de cuatro semanas con la nueva metodología establecida en el 2007, pero anteriormente era de cinco.
5. **Desempleo:** A partir del año 2007, del cálculo del desempleo oculto se excluye la categoría “no busca trabajo por OPOSICIÓN FAMILIAR O DEL CÓNNYUGE”. Además de esto, en el cálculo del desempleo abierto no se considera la categoría “esta a la espera de un trabajo”, pues esta categoría a partir del 2007 es parte del desempleo oculto.

Los cambios en la metodología mencionados anteriormente hacen necesario realizar un quiebre en el año 2007 para el análisis del empleo, como ya fue mencionado.

También es importante destacar que el mercado laboral presenta un claro componente estacional, lo que implica que existen meses en el año donde el mercado laboral sufre un shock temporal positivo. Un claro ejemplo de esto son las épocas navideñas (mediados del mes de noviembre hasta mediados del mes de enero) en donde los sectores productivos se abastecen de mano de obra temporal para poder satisfacer este incremento transitorio de la demanda.

Principales definiciones.

Población Económicamente Activa: Son económicamente activas las personas en edad de trabajar que:

- Trabajaron al menos una hora durante el período de referencia de la medición (por lo general, la semana anterior) en tareas con o sin remuneración, incluyendo la ayuda a otros miembros del hogar en alguna actividad productiva o en un negocio o finca del hogar
- Si bien no trabajaron, tenían algún empleo o negocio del cual estuvieron ausentes por enfermedad, huelga, licencia, vacaciones u otras causas; y
- No comprendidas en los dos grupos anteriores, que estaban en disponibilidad de trabajar. Se excluyen las personas que se dedican solo al quehacer doméstico o solo a estudiar, así como a los que son solo pensionistas y a los impedidos de trabajar por invalidez, jubilación, etc.

Tasa de Desempleo: tasa entre el número de desocupados y la población económicamente activa. Por desocupados/as se entiende a aquellas personas en edad de trabajar que durante el período de referencia de la medición (la última semana) no tenían empleo y estaban disponibles para trabajar. Abarca tanto a:

- Aquellos trabajadores/as que se quedaron sin empleo por despido o renuncia (cesantes), y
- A quienes se incorporan por primera vez al mercado de trabajo (trabajadores nuevos). Se refiere solo a la población económicamente activa (PEA).

Tasa de Subempleo: Debido a las diferencias entre las metodologías, no se puede detallar un solo concepto de subempleo para todo el periodo estudiado.

- **Antes del 2007:** Número de personas en edad de trabajar subempleadas en el sector moderno sumadas al número de empleados/as en el sector informal, expresado como porcentaje del total de la población económicamente activa (PEA) en un determinado año.
- Se consideran subempleados/as las personas que involuntariamente trabajan menos de 40 horas a la semana en el sector moderno (subempleo visible) o quienes, aunque trabajen 40 horas o más en el sector moderno, reciben un ingreso por debajo del salario mínimo vital (subempleo invisible).
- Un tercer concepto es el de subempleo productivo, que tiene que ver con la forma de organización y las características de las unidades productivas y que tiene correspondencia con la definición de sector informal. La suma de los tres conceptos anteriores es un indicador del nivel bruto de subempleo en la economía. El sector moderno del mercado laboral está compuesto por los ocupados/as que trabajan en establecimientos económicos: con más de 5 trabajadores/as; de hasta 5 personas, pero como patronos, cuentapropistas y asalariados profesionales o técnicos; y de hasta 5 personas en cuya rama de actividad no haya establecimientos que puedan ser clasificados como informales. El sector informal está integrado por los ocupados/as por cuenta propia, trabajadores/as familiares no remunerados, patronos y asalariados/as de establecimientos de hasta 5 trabajadores, con excepción de quienes desarrollan actividades de nivel profesional o técnico.
- **Desde el 2007:** se expresa como el porcentaje de la PEA que representan los individuos subempleados, ya sea por insuficiencia de horas (trabajan menos de 40 horas a la semana de manera involuntaria) o por:
 - Trabajan 40 horas o más, tienen ingresos iguales o superiores al salario unificado legal y están dispuestos y disponibles a trabajar más horas o cambiarse de trabajo para trabajar más horas.
 - Personas ocupadas que trabajan 40 horas o más y tienen ingresos menores al salario unificado legal y están dispuestas y disponibles a trabajar más horas.

- Personas ocupadas que trabajan menos de 40 horas, tienen ingresos menores al salario unificado legal y no están dispuestos o disponibles a trabajar.
- **Tasa de Ocupación Global:** Proporción de la PEA que representan las personas en edad de trabajar que se encuentran ocupadas. Por ocupadas se entiende aquellas personas que trabajaron al menos una hora en el período de referencia de la medición (semana anterior), o aunque no trabajaron, tuvieron algún trabajo del cual estuvieron ausentes por motivos tales como vacaciones, enfermedad, licencia por estudios, etc. Se considera ocupadas también a aquellas personas que realizan actividades productivas dentro del hogar, aunque las actividades desarrolladas no guarden las formas típicas del trabajo asalariado e independiente. Se incluyen también como ocupados a los trabajadores familiares no remunerados que trabajan periódicamente en el negocio familiar.

2.1.2 Homogenización de la base de datos.

Dado que el programa diseñado opera sobre una base de datos maestra, primero fue necesario decidir que variables se van a utilizar para la estimación de los indicadores.

Como fue mencionado, existe un importante quiebre en la metodología a partir de junio del 2007, por ello algunas de las variables que se utilizan para el cálculo de los indicadores bajo la nueva metodología no existen en la antigua y viceversa, mientras que la mayoría de las variables que se mantienen a lo largo de todo el periodo estudiado, ya no son estadísticamente comparables.

Para el diseño de la base de datos maestra se utilizaron la lista de todas las variables que están involucradas en el cálculo de los indicadores². Estas variables pueden no existir en un determinado trimestre por 2 razones. Primero, en el caso de que estas variables no existan ya que no estaban definidas en la metodología utilizada por el INEC, se procedió a crearlas con *missing values* para no alterar el esquema de la base.

El segundo caso se da cuando las variables no existen ya que el INEC decidió en ese trimestre agregarlas en otras variables categóricas. Por ejemplo, en ciertos trimestres se utiliza la variable *ocupa* que toma el valor de 1 si la observación se clasifica como ocupada, pero en otros trimestres esta y otras variables del estado laboral del encuestado se encuentran agregadas en la variable *conduct*, que toma

² Para una lista de los indicadores estimados y una explicación del método de cálculo, véase anexo I.

un valor diferente según la condición de actividad del individuo (ocupado pleno, subempleado, desempleado, etc.).

Nótese que en el ejemplo mencionado, la variable *condact* no incluye una categoría para los ocupados, sino para su sub-clasificación (ya que un ocupado puede encontrarse plenamente empleado o subempleado). Estas variables categóricas fueron procesadas para obtener las variables como son requeridas.

Una vez realizado el procesamiento de las bases cada trimestral para la construcción de la base maestra, los nombres de las variables seleccionadas fueron homogenizados para su estudio a lo largo del tiempo.

Lamentablemente los quiebres en la metodología mencionados no fueron los únicos que se realizaron en el periodo de análisis. Además de esto para la mayoría de las variables analizadas, se realizaron varios cambios en las opciones de las respuestas permitidas para el encuestado. Por esto fue necesario analizar las variables utilizadas para asegurar que las respuestas de las mismas sean comparables a lo largo del tiempo. Si el caso lo permitía, se re-categorizó las variables para mantener la consistencia de la base.

Las variables utilizadas para posibles desagregaciones para los indicadores (solo 2 pueden ser seleccionadas de forma simultánea) se describen en la tabla 1:

Tabla 1: Variables escogidas para desagregación de indicadores sociales

Variable	Label creada	Descripción
Grupos de Edad	edad_group	Los intervalos de edad utilizados son los recomendados por la CEPAL ³ . Cabe recalcar que en las encuestas todas las personas mayores a 98 años de edad son agrupadas en una sola categoría, donde la variable edad toma el valor de 98. Dado que esto le daría un peso diferente a estas observaciones en comparación a las otras edades, estas observaciones fueron eliminadas de la base.
Etnia	etnia	Variable procesada de la pregunta <i>¿Cómo se considera?</i> Las categorías estandarizadas para todo el periodo son: blanco, indígena, mestizo, negro y no definidos. Es importante destacar que esta pregunta no refleja verdaderamente la etnia del encuestado, sino su percepción de la misma.
Nivel de Instrucción	nivinst	Variable creada en base a la pregunta <i>¿Cuál es el nivel de instrucción y año más alto que aprobó?</i> La

³ Los intervalos de edad utilizados son: 10-17, 18-29, 30-39, 40-49, 50-64 y Mayores de 65 años.

		forma de clasificación de la educación formativa sufrió un cambio, de primaria y secundaria a educación básica y educación media, el cual fue realizado paulatinamente desde principios de la última década. Así, esta pregunta fue diseñada para que el encuestado conteste según el tipo de formación que recibió, a pesar de la equivalencia que existe, por ejemplo, entre la primaria y los primeros 3 años de la secundaria con la educación básica. Las respuestas fueron estandarizadas en las siguientes categorías: Centro de Alfabetización, Educación Básica, Educación Media, Educación Superior, Post Grado, Ninguno y No contesta.
Región Natural	<i>region_natural</i>	En ciertos trimestres además de las 3 principales regiones naturales del Ecuador (la región insular no es incluida en el diseño del marco muestral), se incluyó una cuarta categoría, “Costa no delimitada”, pero esta fue agregada a la categoría de costa para mantener el estándar
10 ciudades más importantes, y resto país	<i>ciudad_ind</i>	Construida a partir de la variable <i>ciudad</i> incluida en todas las bases de datos. A partir del código utilizado por el INEC para clasificar los centros poblados se identificaron las 10 ciudades más importantes, el restante de las diferentes parroquias fueron agregadas bajo la categoría “Resto País”.

Para el cálculo de los indicadores se homogenizaron las siguientes variables descritas en la tabla 2:

Tabla 2: Variables a homogenizar para el cálculo de indicadores

Variable	Label creada	Descripción
Años de escolarización	<i>anosaprob</i>	Variable creada también en función de la misma pregunta utilizada para obtener el Nivel de Instrucción. Para esto se trabajó con la duración estándar de cada uno de los niveles de educación mencionados.
Ingresos	<i>ingrl</i>	Se utilizó como base la variable de ingresos procesada por el INEC. Es más adecuado denominar a esta variable como una medida de beneficio reportado de las actividades laborales del encuestado en lugar que de ingresos, ya que es posible obtener valores negativos. Esta es una de las variables que sufrió importantes cambios en su metodología de cálculo a partir del año 2007. Además de esto, los ingresos negativos

		<p>fueron agregados por el INEC en una sola categoría a partir del 2007, a diferencia de lo que ocurría antes donde el valor de ingresos (sin importar si este fuese positivo o negativo) era reportado.</p> <p>Todos estos ingresos negativos fueron agregados al 0 para no perder el impacto negativo que debiesen tener sobre la media. Pero debido a esto es necesario reconocer que la media aquí reportada esta sesgada a la alza. Por esto también se incluyo un indicador del porcentaje de personas que reportaron un ingreso negativo del total de personas que reportaron ingresos.</p>
--	--	--

Las variables utilizadas para la construcción de los diferentes indicadores del mercado laboral fueron las ya procesadas por el INEC en base a las diferentes metodologías mencionadas. Estos indicadores fueron convertidos en variables dicotómicas para su utilización directa como variable dependiente en el modelo econométrico explicado en la siguiente sección.

2.2 Modelo Econométrico

El objetivo del observatorio es ofrecer una herramienta versátil para la estimación de indicadores del mercado laboral a distintos niveles de desagregación y para diferentes periodos y para ello el principal insumo para la consecución de este objetivo es la ENEMDU.

En la metodología del diseño de la ENEMDU para diciembre del 2010 se menciona que el marco muestral para esta es construido por el INEC como “...una muestra probabilística y trietápica”. Las unidades muestrales en las diferentes etapas del muestreo son seleccionadas asegurando “grupos homogéneos al interior y, heterogéneos entre sí... para aumentar la eficiencia de la muestra maestra.” En el nivel más bajo de las etapas mencionadas se realiza un muestreo aleatorio simple. Pero dado que el objetivo de la plataforma es el permitir el cálculo de indicadores a diferentes niveles de desagregación, el referenciar la submuestra resultante con las unidades muestrales especificadas por el INEC sería muy pesado en términos computacionales. Sin embargo, si esto no es referenciado correctamente se estaría incurriendo en heteroscedasticidad por agrupación⁴.

Por esto fue necesario el diseñar un método econométrico que permita una estimación de los indicadores planteados que sea robusta a cualquier estructura de heteroscedasticidad producida por las desagregaciones realizadas.

⁴Greene, William. *Econometric Analysis*, Fourth Edition. Pag 505.

Como los indicadores propuestos son medias de diferentes variables, se plantea una regresión de la variable de interés con respecto solamente a una constante. Formalmente,

$$y_{ij} = \beta_i + \mu_{ij}; \quad \mu_{ij} \sim N(0, \sigma_{ij}^2) \quad (1)$$

Donde y_{ij} representa el valor de la variable de interés para el individuo j de la desagregación i , β_i representa la media de la variable de interés bajo la desagregación i , y μ_{ij} es el error. Nótese que la varianza del error (σ_{ij}) se encuentra indexada tanto a i como a j , permitiendo la presencia de heteroscedasticidad según el nivel y tipo de desagregación escogida por el usuario. Sin pérdida de generalidad se puede expresar la ecuación (1) de forma matricial, como se muestra a continuación,

$$Y_i = X_i \beta_i + U_i \quad (2)$$

$$Y_i = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_i} \end{bmatrix}_{n_i \times 1}; \quad X_i = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n_i \times 1}; \quad U_i = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{n_i} \end{bmatrix}_{n_i \times 1}$$

Donde $U_i \sim N(0, \Omega_i)$. Y se define a la matriz de varianzas y covarianzas de los errores (Ω_i) de la siguiente manera:

$$\Omega_i = \begin{bmatrix} \varrho_1 & 0 & \dots & 0 \\ 0 & \varrho_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \varrho_{n_i} \end{bmatrix} \sigma_i^2$$

Así como en (1), todas las matrices están indexadas por i para indicar que el tamaño de la muestra utilizada para el cálculo de cada indicador depende de las desagregaciones que determine el usuario en la plataforma.

Dado que, como fue mencionado, el marco muestral de la ENEMDU fue diseñado mediante un muestreo trietápico; es necesario reconocer que cada observación no tiene la misma probabilidad de selección y por ende el mismo peso en la muestra (exp). Es por esto que para la estimación de β_i se minimizará la suma de los cuadrados de los errores, ponderándolos por su respectivo peso en la muestra. Formalmente,

$$Min_{\beta_i} \sum_{j=1}^{n_i} exp_j \mu_j^2 = U_i' F_i U_i \quad (3)$$

$$F_i = \begin{bmatrix} \text{fexp}_1 & 0 & \dots & 0 \\ 0 & \text{fexp}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{fexp}_{n_i} \end{bmatrix}$$

Resolviendo esta multiplicación se obtiene,

$$U_i' F_i U_i = (Y_i - X_i \beta_i)' F_i (Y_i - X_i \beta_i) = Y_i' F_i Y_i - 2\beta_i' X_i' F_i Y_i + \beta_i' X_i' F_i X_i \beta_i$$

Y minimizando con respecto a β_i ,

$$\frac{\delta U_i' F_i U_i}{\delta \beta_i} = -2X_i' F_i Y_i + 2X_i' F_i X_i \beta_i = 0$$

$$\hat{\beta}_i = (X_i' F_i X_i)^{-1} X_i' F_i Y_i \quad (4)$$

Es posible resolver esta expresión recordando que X_i es un vector columna de unos, por lo que se obtiene:

$$\hat{\beta}_i = \left(\sum_{j=1}^{n_i} \text{fexp}_j \right)^{-1} \sum_{j=1}^{n_i} \text{fexp}_j y_j$$

$$\hat{\beta}_i = \frac{\sum_{j=1}^{n_i} \text{fexp}_j y_j}{\sum_{j=1}^{n_i} \text{fexp}_j} \quad (5)$$

Como se puede ver, esta es la expresión de la media ponderada que se esperaría de datos con diferentes pesos en la muestra general. En base a esto es simple demostrar que el estimador planteado es insesgado a pesar de la existencia de heteroscedasticidad. Pero para realizar inferencia no solo es importante conocer el valor estimado (e insesgado) del indicador, sino también tener una medida consistente de su varianza. Para esto se obtiene una expresión del segundo momento poblacional del estimador.

$$V(\hat{\beta}_i) = E[(\hat{\beta}_i - \beta_i)(\hat{\beta}_i - \beta_i)']$$

$$(\hat{\beta}_i - \beta_i) = (X_i' F_i X_i)^{-1} X_i' F_i (X_i \beta_i + U_i) - \beta_i = (X_i' F_i X_i)^{-1} X_i' F_i U_i$$

$$V(\hat{\beta}_i) = E[(X_i' F_i X_i)^{-1} X_i' F_i U_i U_i' F_i X_i (X_i' F_i X_i)^{-1}]$$

Como se asume que las matrices X_i y F_i son exógenas, pueden salir de la esperanza,

$$V(\hat{\beta}_i) = (X_i' F_i X_i)^{-1} X_i' F_i E[U_i U_i'] F_i X_i (X_i' F_i X_i)^{-1} \quad (6)$$

Donde

$$E[U_i U_i'] = \Omega_i$$

Es importante recalcar que esta expresión difiere de la obtenida mediante la estimación convencional por Mínimos Cuadrados Ordinarios Ponderados (MCOP), donde se asume que los pesos son una medida de la heteroscedasticidad presente en el modelo ($exp_j = 1/\rho_j$). Si esto fuese cierto, la expresión (6) podría reducirse de tal forma que se obtenga $V(\hat{\beta}_i) = (X_i' F_i X_i)^{-1} \sigma_i^2$ (estimador de la varianza incluido generalmente en los diferentes paquetes estadísticos para MCOP). Pero como fue indicado, dado que no se está considerando todas las etapas del muestreo al momento de realizar la desagregación, no es posible asumir esta conveniente relación entre los pesos y la heteroscedasticidad. Esto también implica que no es posible utilizar el método de weighted regresión (MCOP) incluido en la mayoría de paquetes estadísticos ya que el estimador de la varianza que estos presenten se encontraría sesgado.

A pesar de esto, en base a White (1980), se puede desarrollar un estimador consistente de esta varianza.

Para esto se define,

$$\Sigma_i = \frac{1}{n_i} X_i' F_i \Omega_i F_i X_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \rho_{ij} \sigma_i^2 \overrightarrow{exp_j} \overrightarrow{exp_j}'$$

Donde $\overrightarrow{x_{ij}}$ representa a la columna j de la matriz X_i .

Se puede demostrar bajo condiciones muy generales que

$$S_{i0} = \frac{1}{n_i} \sum_{j=1}^{n_i} \widehat{\mu}_{ij}^2 \overrightarrow{exp_j} \overrightarrow{exp_j}' \quad (7)$$

Es un estimador consistente de Σ_i .

Reemplazando (7) en (6) y multiplicando por n_i para no afectar la igualdad se obtiene,

$$V(\hat{\beta}_i) = n_i (X_i' F_i X_i)^{-1} S_{i0} (X_i' F_i X_i)^{-1} \quad (8)$$

Como fue realizado para obtener (5), se puede resolver la expresión (8) para obtener,

$$V(\hat{\beta}_i) = n_i \left(\sum_{j=1}^{n_i} exp_j \right)^{-1} S_{i0} \left(\sum_{j=1}^{n_i} exp_j \right)^{-1}$$

Y dado que X_i es un vector columna de unos, se conoce que $\overrightarrow{x_{ij}} = 1$, por lo que se puede expresar (7) como,

$$S_{io} = \frac{1}{n_i} \sum_{j=1}^{n_i} \widehat{\mu}_{ij}^2 \text{exp}_j^2(1)^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \widehat{\beta}_i)^2 \text{exp}_j^2$$

Reemplazando esta expresión en la previamente obtenida para la varianza del estimador se obtiene,

$$V(\widehat{\beta}_i) = \frac{\sum_{j=1}^{n_i} (y_{ij} - \widehat{\beta}_i)^2 \text{exp}_j^2}{(\sum_{j=1}^{n_i} \text{exp}_j)^2} \quad (9)$$

Como ya fue indicado, esta metodología permite la estimación de indicadores robustos a cualquier estructura de heteroscedasticidad que pueda ser ocasionada por las diferentes desagregaciones que realice el usuario de la plataforma diseñada (a pesar del muestreo trietápico realizado por el INEC).

2.3 Diseño del sistema

El sistema está diseñado desde cero e implementado por completo en C# 3.5 con SQLite 3, como una plataforma de consulta de escritorio. No obstante, desde el principio se ha mantenido como enfoque facilitar la portabilidad del mismo hacia ambientes multiusuario orientados al Internet. Para ello, el mismo ha sido implementado en varias capas, con un enfoque orientado a objetos. Mediante ello se busca la portabilidad y reusabilidad del código desarrollado, de tal forma que su integración en ambientes orientados a servicios XML y/o arquitecturas ASP.NET sea factible.

Con esta filosofía en mente, se desarrollaron cuatro ensamblados distintos en donde se encapsulan por separado las capas de servicios de consulta, cálculo y lógica de negocios e interfaces de usuario y reporte de datos. De esta forma, permite reutilizar el código al máximo al momento de portar el sistema a un entorno multiusuario. La siguiente lista detalla los componentes que se desarrollaron como parte de esta plataforma:

- *NumericMethodsLib.dll* – Este componente implementa la lógica de abstracción de los métodos numéricos utilizados en el sistema para resolver el modelo econométrico planteado en la sección III. Define el namespace (espacio de nombres) *NumericMethodsLib*.
- *DataServices.dll* – Define la interface genérica de consulta de datos y la implementación específica en SQLite 3 para el sistema de consulta de escritorio. Este componente debe ser extendido si se piensa portar el sistema con un DataProvider orientado a transacciones multiusuario como SQL Server u Oracle. Define el namespace *DataServices*.

- *IndicatorModel.dll* – En este componente se encapsula los tipos de datos, lógica de negocio y el patrón ViewModel implementado para facilitar la navegación de datos en interfaces de usuario. Define los namespaces ***IndicatorModel.InfoLayer***, ***IndicatorModel.DataModel*** y ***IndicatorModel.ViewModel***.
- *ObservEconFEN.exe* – Ejecutable para Win32 desarrollado en WPF 3.5 con C#. Las interfaces de usuario y lógica de reportes en Excel se implementan en este ensamblado. Define el namespace ***ObservEconFEN***.

La descripción del sistema se hará enfocándose en cada uno de los espacios de nombres más importantes por separado. Para ello se presentara un diagrama de clases y su descripción, a la par describir las funciones generales de las clases implementadas en cada uno de ellos.

NumericMethodsLib

La figura 1 muestra el diagrama de clase de los componentes de este espacio de nombres. El componente implementa la clase *Matrix*, la cual encapsula los algoritmos de métodos numéricos necesarios para las operaciones matriciales más comúnmente utilizadas en los algoritmos de estimación econométricos, tales como la suma, el producto, el producto punto, inversión de matrices, descomposición espectral, entre otras operaciones. Esto le permitirá al código que consume este espacio de nombres el abstraer este tipo de operaciones y tratar directamente con variables de tipo *Matrix* y sus operaciones asociadas, de forma similar a otros productos tales como MATLAB, entre otros. También se definen las excepciones que maneja la clase *Matrix* para estandarizar el manejo de errores.

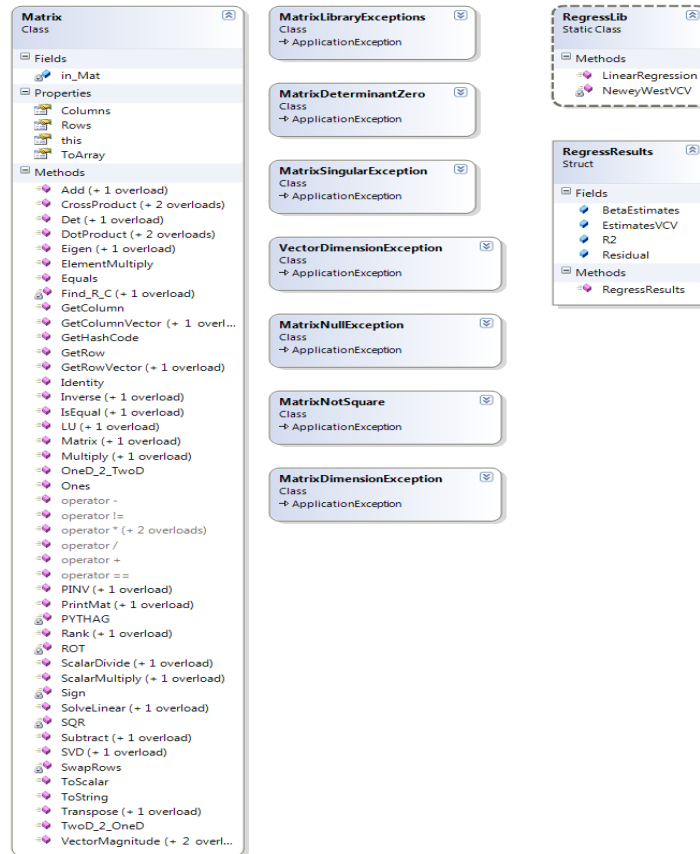


Figura 1: Diagrama de clases para el namespace *NumericMethodsLib*.

Adicionalmente, se implementa la clase *RegressLib* aprovechando la implementación del tipo *Matrix* para implementar algoritmos de regresión lineal por mínimos cuadrados y estimaciones de varianza homocedástica y corregida por heterocedasticidad y autocorrelación por el método de Newey-West.

DataServices

El diseño de esta capa esta pensado para facilitar la integración de las capas superiores de la plataforma a entornos diferentes al de una aplicación de escritorio en Win32.

Para ello, la figura 2 muestra la definición de la interface *IDbCnx* y la estructura *DbParam* que permiten de forma genérica manejar el acceso a la base de datos necesario para la aplicación. No obstante, solo se implementa la interface mediante la clase *SQLiteDbCnx*, la cual da soporte al motor de base de datos SQLite 3. Este es un

sistema de gestión de bases de datos relacional compatible con ACID, contenida en una relativamente pequeña (~275 KB) biblioteca de dominio público escrita en C.

La biblioteca SQLite se enlaza con la plataforma pasando a ser parte integral del mismo. Este componente aprovecha la funcionalidad de SQLite a través de llamadas simples a subrutinas y funciones. Esto reduce la latencia en el acceso a la base de datos, debido a que las llamadas a funciones son más eficientes que la comunicación entre procesos.

El conjunto de la base de datos (definiciones, tablas, índices, y los propios datos), son guardados como un sólo fichero estándar en la máquina host. Este diseño simple se logra bloqueando todo el fichero de base de datos al principio de cada transacción. No obstante, para migrar la plataforma a un entorno multiusuario, se debe implementar la interface *IDbCnx* con una clase para otros proveedores de datos transaccionales como *SQLServer*.

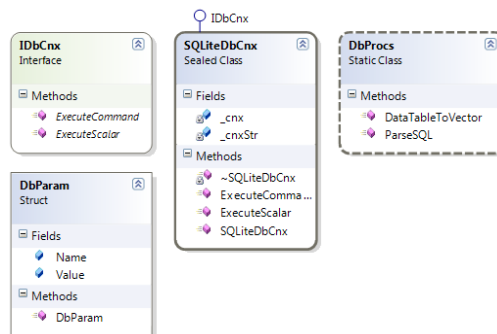


Figura 2: Diagrama de clases para el namespace *DataServices*.

IndicatorModel.InfoLayer

En este espacio de nombres se definen los tipos de datos que la aplicación utiliza para diagramar la lógica de negocios que se implementa en el namespace *IndicatorModel.DataModel*. La figura 3 enumera todos los tipos de datos utilizados, entre ellos la definición de los conceptos de trimestre, punto de datos, serie de tiempo, indicador, definición de indicador, variable de agregación, nivel de agregación, entre otros.

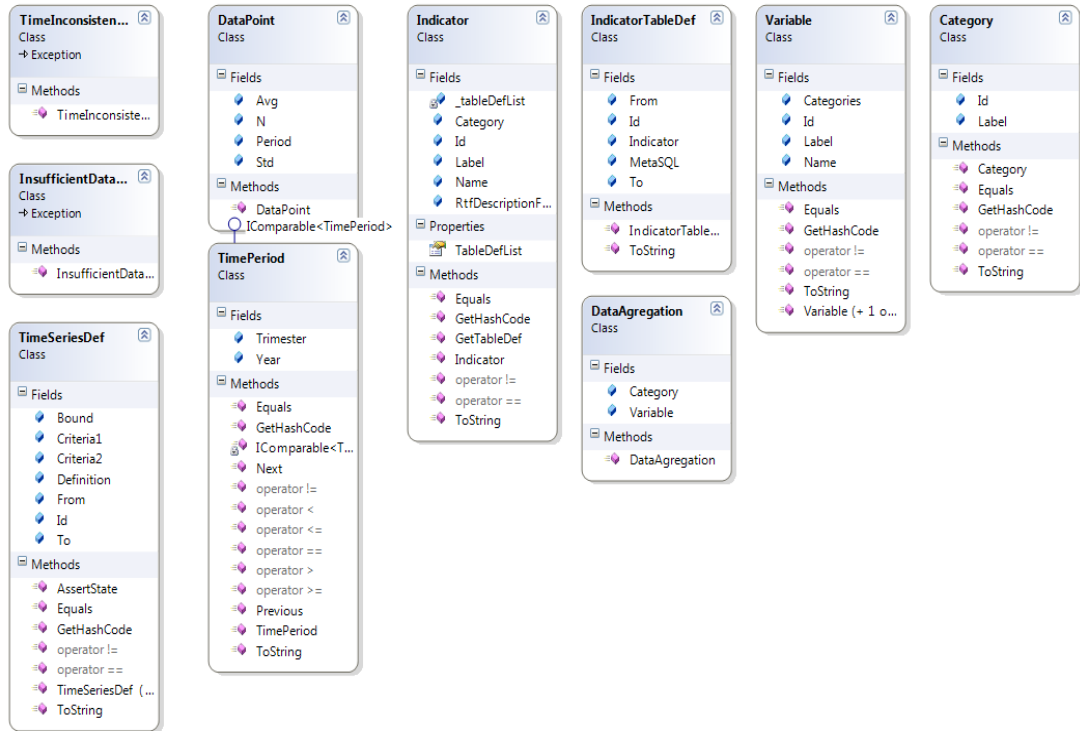


Figura 3: Diagrama de clases para el namespace *IndicatorModel.InfoLayer*.

IndicatorModel.DataModel

En este componente se definen las clases que implementan la lógica de negocio de la plataforma. La figura 4 muestra el diagrama de clases de este namespace, las cuales agrupan la funcionalidad que soporta la aplicación. Cada de una de ellas se encarga de tareas distintas dentro de la construcción de las colecciones de objetos con las cuales se construyen las tablas de series de tiempo de los indicadores resultantes del sistema.

La clase *ObjectModelBuilder* es la que se encarga de interactuar directamente con el namespace *DataServices*, aprovechando sus interfaces genéricas para mantener una conexión a la base datos independiente del motor utilizado e además de implementar procedimientos para construir las listas y colecciones indexadas de los objetos que contienen la información necesaria para representar los diversos indicadores, sus definiciones y la data necesaria para la definición de las series de tiempo en la plataforma.

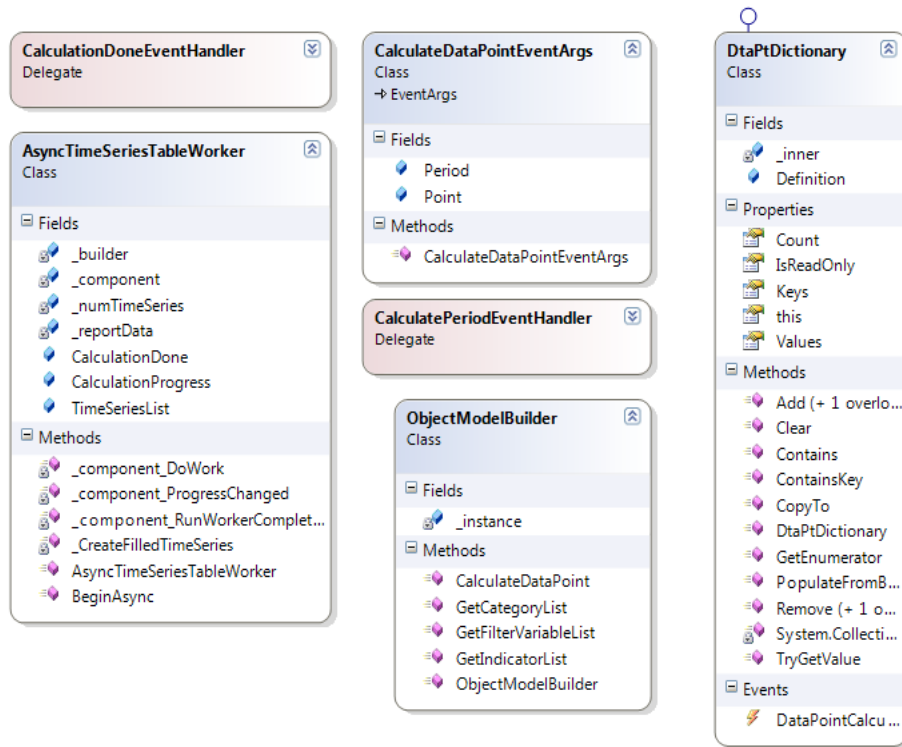


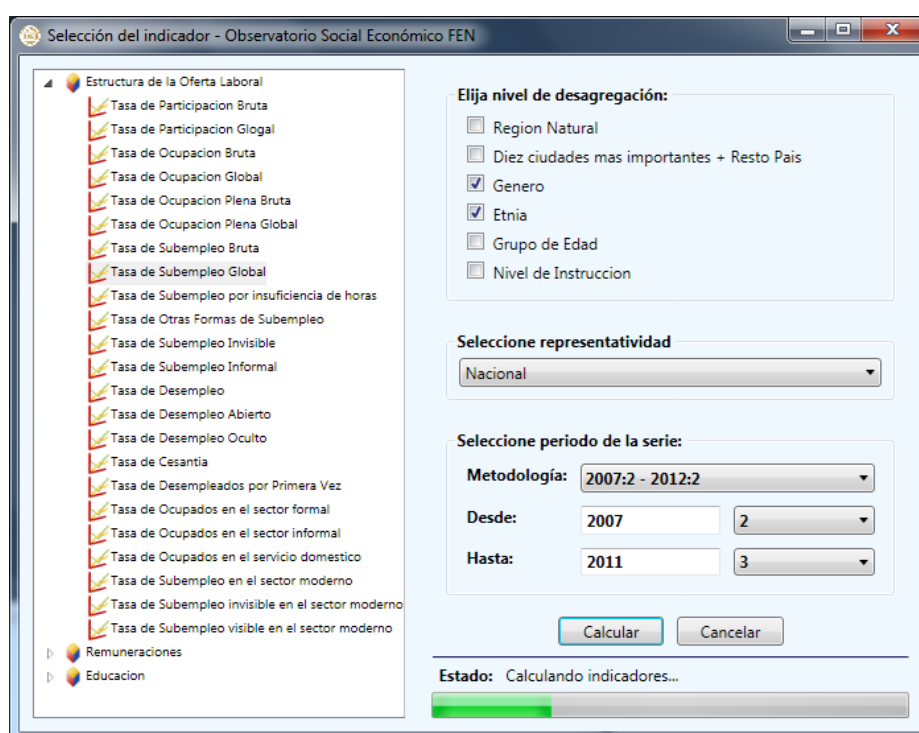
Figura 4: Diagrama de clases para el namespace *IndicatorModel.DataModel*.

En este namespace también se define la clase *DtaPtDictionary*, la cual deriva la funcionalidad de un diccionario genérico básico, pero sobrescribe sus métodos de poblado de datos para permitirle ser llenado asincrónicamente a partir de una instancia previa de la clase *ObjectModelBuilder*. Adicionalmente, implementa la clase *AsyncTimeSeriesTableWorker*, el cual implemente un modelo asincrónico basado en el modelo *ViewModel* implementado en este namespace para construir asincrónicamente las tablas de series de tiempo que posteriormente serán la salida final de la plataforma de consulta. Cabe mencionar que este componente implementa sus procesos multi-hilo mediante en modelo por apartamentos uni-proceso. Es decir, cada subproceso recibe información de entrada y comunica sus resultados finales al proceso principal (AppDomain del ejecutable *ObservEconFEN.exe*). Esto se utiliza para mantener la interface de usuario respondiendo mientras el sistema procede con los cálculos econométricos que esta capa efectúa para construir la series de tiempo individuales que forman parte del reporte final. Finalmente, se definen en este namespace las clases de soporte para definir los eventos que forman parte del modelo asincrónico descrito.

3. Resultados

El proceso detallado en las secciones anteriores da como resultado final un sistema de información o plataforma de indicadores socio – económicos a ser consultados en diferentes momentos del tiempo a partir del año 2003, por las razones explicadas a lo largo del documento.

Al ejecutar dicha plataforma, aparece la siguiente aplicación representada en la figura 5.



Así por ejemplo, se tiene que si el usuario desea conocer cuál es la tasa de desempleo global, es decir, el porcentaje de la población económicamente activa que desea trabajar más horas o que recibe un salario inferior al salario mínimo vital, a la vez que se requiere dicho indicador desagregado por género y étnia para el periodo que parte del II trimestre del 2007 hasta el III trimestre del 2011, si tiene que escoger las opciones mostradas en las figura 5, para luego escoger la opción de “calcular”.

Los resultados que arroja el sistema son presentados en una hoja de Excel, como se muestra en la tabla 3.

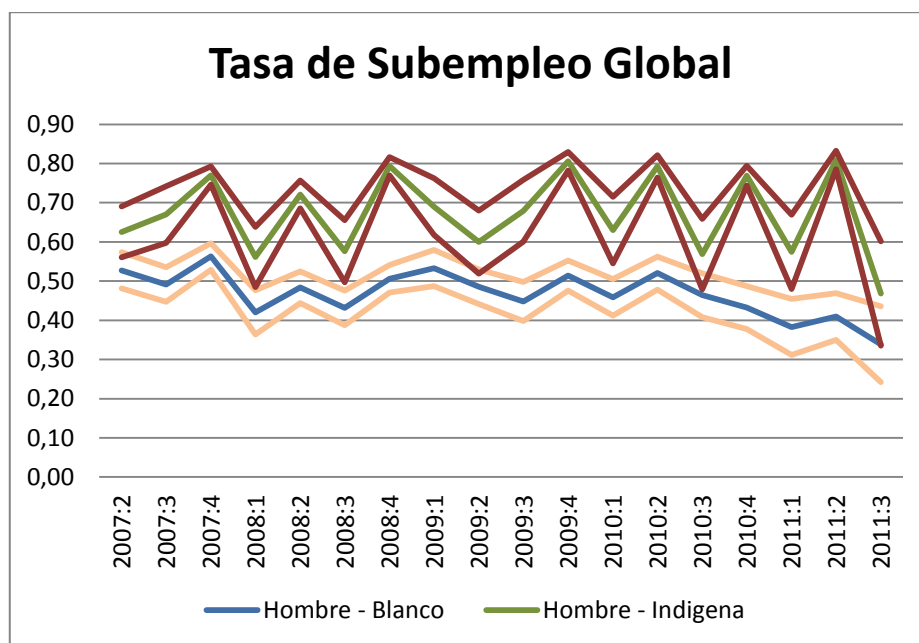
Tabla 3: Resultados de la consulta Tasa de subempleo global desagregada por género masculino y etnia.

Periodo	Hombre Indic.	Blanco Std.	Hombre Indic.	Indigena Std.	Hombre Indic.	Mestizo Std.	Hombre Indic.	Negro Std.	Hombre Indic.	Otro Std.
2007:2	0,53	0,02	0,63	0,03	0,53	0,01	0,63	0,04	0,62	0,04
2007:3	0,49	0,02	0,67	0,04	0,51	0,01	0,53	0,04	0,52	0,03
2007:4	0,56	0,02	0,77	0,01	0,58	0,01	0,60	0,03	0,65	0,03
2008:1	0,42	0,03	0,56	0,04	0,51	0,01	0,47	0,05	0,61	0,05
2008:2	0,48	0,02	0,72	0,02	0,56	0,01	0,61	0,03	0,59	0,03
2008:3	0,43	0,02	0,58	0,04	0,50	0,01	0,54	0,04	0,60	0,04
2008:4	0,51	0,02	0,79	0,01	0,55	0,01	0,62	0,03	0,578	0,03
2009:1	0,53	0,02	0,69	0,04	0,51	0,01	0,62	0,04	0,61	0,05
2009:2	0,49	0,02	0,60	0,04	0,50	0,01	0,58	0,05	0,63	0,04
2009:3	0,45	0,03	0,68	0,04	0,52	0,01	0,58	0,05	0,61	0,05
2009:4	0,51	0,02	0,81	0,01	0,59	0,00	0,64	0,03	0,61	0,03
2010:1	0,46	0,02	0,63	0,04	0,52	0,01	0,51	0,05	0,48	0,05
2010:2	0,52	0,02	0,79	0,01	0,57	0,01	0,64	0,02	0,74	0,01
2010:3	0,46	0,03	0,57	0,05	0,48	0,01	0,54	0,05	0,59	0,03
2010:4	0,43	0,03	0,77	0,01	0,53	0,01	0,59	0,02	0,72	0,01
2011:1	0,38	0,04	0,57	0,05	0,48	0,01	0,58	0,04	0,64	0,03
2011:2	0,41	0,03	0,81	0,01	0,53	0,01	0,60	0,02	0,72	0,01
2011:3	0,34	0,05	0,47	0,07	0,43	0,01	0,39	0,05	0,46	0,04

Interpretando los resultados mostrados en la tabla 3, se tiene que aproximadamente el 53% de los hombres que se consideran de raza blanca para el segundo trimestre del 2007, se encuentran subempleados, siendo la desviación estándar de la media 0,02. Haciendo el análisis a lo largo del periodo solicitado, se tiene que este porcentaje cae al 34% para el tercer trimestre del año 2011, demostrando una disminución del subempleo, no solo para los hombres considerados de raza blanca, sino para todas las etnias en general.

Asimismo el sistema arroja estos indicadores para el género femenino, sin embargo, esta es una muestra de los resultados encontrados a partir del procesamiento de la base de datos obtenidos a partir de la encuesta urbana de empleo y desempleo del INEC.

Finalmente el sistema muestra los gráficos respectivos de los indicadores calculados y desagregados según las variables de interés como se muestra en el gráfico 1, donde las bandas representan los intervalos de confianza respectivos, que permiten realizar pruebas de hipótesis.



4. Conclusiones

El Observatorio Social Económico es un sistema de información y consulta de indicadores que permite realizar diagnósticos de la situación de los ecuatorianos en torno a variables como empleo, educación, ingresos, lo que permite a su vez realizar análisis encaminados a la recomendación de políticas públicas.

Se busca así contribuir a fortalecer los vínculos de la ESPOL con la Comunidad en general, desarrollando investigación a partir de la generación de este tipo de herramientas que involucró tanto el desarrollo de modelos econométricos para el procesamiento de la base de datos obtenida a partir de la ENEMDU como de la implementación de herramientas informáticas.

5. Agradecimientos

Los autores desean expresar su agradecimiento a la Facultad de Economía y Negocios de la ESPOL, particularmente a sus directivos, Ph.D Leonardo Estrada y Ec. Mariela Méndez por el apoyo brindado para poder llevar a cabo la implementación del Observatorio Social Económico.

Referencias

Greene, W.: *Econometric Analysis*, cuarta edición, Prentice Hall International, (2000).

White, H.: A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, Journal of the Econometric Society (1980).

Newey, W., West, K.: A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society* (1987).

Sen, A., Kliksberg B.: *Primero la gente*. Ediciones Deusto (2007)

Gamma E., Helm R., Johnson R., Vlissides J, *Design Patterns: Elements of Reusable Object-Oriented Software*

SQLite Documentation web site, <http://www.sqlite.org/docs.html>

Instituto Nacional de Estadística y Censos,
http://inec.gob.ec/estadisticas/?option=com_content&view=article&id=109&Itemid=88

**APENDICE 1:
INDICADORES BASE, CONCEPTOS Y METODOLOGÍA DE
CÁLCULO.**

1. **Tasa de Participación Bruta:** Este indicador muestra el tamaño de la fuerza laboral en relación a la población total. Es decir, mide la proporción de la población total (PT) que se encuentra económicamente activa (PEA) para un año dado. La PEA es la población a partir de 10 años que trabajó al menos 1 hora a la semana; o bien, no habiendo trabajado se encontraba ocupada o en posibilidad de trabajar (desempleada).
2. **Tasa de Participación Global:** Es una medida más precisa que la Tasa de Participación Bruta debido a que al mostrar el tamaño de la fuerza laboral en relación a la población en edad de trabajar (PET) se evitan sesgos demográficos.

Este indicador mide la proporción de la PET (a partir de 10 años) que se encuentra económicamente activa (PEA) en un año dado. La PEA es la población a partir de 10 años que trabajó al menos 1 hora a la semana; o bien, no habiendo trabajado se encontraba ocupada o subempleada.

3. **Tasa de Ocupación Bruta:** Muestra la proporción de la población en edad de trabajar (PET) que se encuentra ocupada en un año dado. Cabe anotar que hasta el 2006 se subdividía a la población ocupada en: ocupados plenos (aquellos que laboran como mínimo la jornada legal, tienen salarios mayores que el salario unificado y no desean trabajar más, o bien laborando menos horas que la jornada legal cumplen con las otras dos condiciones) y subempleados (personas que laboran menos que la jornada legal y desean trabajar más horas, o bien trabajan mínimo la jornada legal pero sus ingresos son menores al salario unificado, entre otras situaciones). A partir del 2007 se agrega la categoría de ocupados no clasificados (aquellos que no entran en las categorías de ocupados plenos u otras formas de subempleo debido a la falta de datos de ingresos u horas laboradas).
4. **Tasa de Ocupación Global:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra ocupada en un año dado. Cabe anotar que hasta el 2006 se subdividía a la población ocupada en: ocupados plenos (aquellos que laboran como mínimo la jornada legal, tienen salarios mayores que el salario unificado y no desean trabajar más, o bien laborando menos horas que la jornada legal cumplen con las otras dos condiciones) y subempleados (personas que laboran menos que la jornada

legal y desean trabajar más horas, o bien trabajan mínimo la jornada legal pero sus ingresos son menores al salario unificado, entre otras situaciones). A partir del 2007 se agrega la categoría de ocupados no clasificados (aquellos que no entran en las categorías de ocupados plenos u otras formas de subempleo debido a la falta de datos de ingresos u horas laboradas).

5. **Tasa de Ocupación Plena Bruta:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra plenamente ocupada en un año dado. La población plenamente ocupada es aquella que labora como mínimo la jornada legal, tiene salarios mayores que el salario unificado y no desea trabajar más, o bien laborando menos horas que la jornada legal cumplen con las otras dos condiciones.
6. **Tasa de Ocupación Plena Global:** Muestra la proporción de la población ocupada que se encuentra plenamente ocupada en un año dado. La población plenamente ocupada es aquella que labora como mínimo la jornada legal, tiene salarios mayores que el salario unificado y no desea trabajar más, o bien laborando menos horas que la jornada legal cumplen con las otras dos condiciones.
7. **Tasa de Subempleo Bruta:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra subempleada en un año dado. La población subempleada es aquella que labora menos que la jornada legal y desea trabajar más horas, o bien trabaja mínimo la jornada legal pero sus ingresos son menores al salario unificado, entre otras situaciones).
8. **Tasa de Subempleo Global:** Muestra la proporción de la población ocupada que se encuentra subempleada en un año dado. La población subempleada es aquella que labora menos que la jornada legal y desea trabajar más horas, o bien trabaja mínimo la jornada legal pero sus ingresos son menores al salario unificado, entre otras situaciones).
9. **Tasa de Subempleo por Insuficiencia de Horas:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra subempleada por insuficiencia de horas (subempleo visible) en un año dado. El subempleo visible es aquel en el cual las personas laboran menos que la jornada legal y desean o están dispuestos a trabajar más horas.

10. **Tasa de Subempleo por Otras formas de Subempleo:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra subempleada por otras formas de subempleo en un año dado. Esto se da cuando las personas laborando la jornada legal o incluso más horas y recibiendo salarios superiores al salario mínimo, aún desean trabajar más horas o cuando trabajan menos que la jornada legal, reciben salarios inferiores al mínimo y no están dispuestas a trabajar más horas.
11. **Tasa de Subempleo Invisible:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra en subempleo invisible en un año dado. El subempleo invisible es aquel en el cual las personas ganan menos que el salario mínimo.
12. **Tasa de Subempleo Informal:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra en subempleo informal en un año dado. Al subempleo informal el INEC lo cataloga en función de poseer o no RUC o registros contables y en función del tamaño del establecimiento.
13. **Tasa de Desempleo:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra desempleada en un año dado. Desde el 2007 con la nueva metodología aplicada, el desempleo es la suma del desempleo abierto (sin empleo y buscando uno) y el desempleo oculto (sin empleo, no busca uno pero si está disponible para trabajar).
14. **Tasa de Desempleo Abierto:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra en desempleo abierto en un año dado. Se considera desempleo abierto cuando los individuos están sin empleo pero buscan tener uno.
15. **Tasa de Desempleo Oculto:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra en desempleo oculto en un año dado. Se considera desempleo oculto cuando los individuos están sin empleo, no buscan tener uno por algunos motivos, ya sea pero si están disponibles para trabajar.
16. **Tasa de Cesantía:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra cesante en un año dado. Se considera cesante

a toda persona que se encuentra desempleada pero que ha trabajado anteriormente.

17. **Tasa de Desempleados por Primera Vez:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra desempleada por primera vez en un año dado. Se considera desempleado por primera vez a toda persona que se encuentra desempleada pero que no ha trabajado anteriormente.
18. **Tasa de Ocupados en el Sector Formal:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra ocupada en el sector formal en un año dado. Éste es aquel que cuenta con RUC o registros contables.
19. **Tasa de Ocupados en el Sector Informal:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra ocupada en el sector informal en un año dado. El sector informal es aquel que no cuenta con RUC o registros contables.
20. **Tasa de Ocupados en el Servicio Doméstico:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra ocupada en el servicio doméstico en un año dado.
21. **Tasa de Subempleo en el Sector Moderno:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra subempleada en el sector moderno en un año dado. Se define como sector moderno aquel que incorpora el uso de tecnología.
22. **Tasa de Subempleo Invisible en el Sector Moderno:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra en subempleo invisible en el sector moderno en un año dado. Se define como sector moderno aquel que incorpora el uso de tecnología. El subempleo invisible es aquel en el cual las personas ganan menos que el salario mínimo.
23. **Tasa de Subempleo Visible en el Sector Moderno:** Muestra la proporción de la población económicamente activa (PEA) que se encuentra en subempleo visible en el sector moderno en un año dado. Se define como sector moderno aquel que incorpora el uso de tecnología. El subempleo

visible es aquel en el cual las personas laboran menos que la jornada legal y desean o están dispuestos a trabajar más horas.

24. **Ingreso Laboral Personal Promedio:** Es el ingreso promedio de todas las personas que perciben ingresos y están ocupadas en un año dado.

25. **Promedio de Años de Escolaridad Terminados de la PEA:** Muestra el promedio de años de escolaridad de la población económicamente activa (PEA) en un año dado.

26. **Promedio de Años de Escolaridad Terminados de los Ocupados:** Muestra el promedio de años de escolaridad de la población ocupada en un año dado. Cabe anotar que hasta el 2006 se subdividía a la población ocupada en: ocupados plenos (aquellos que laboran como mínimo la jornada legal, tienen salarios mayores que el salario unificado y no desean trabajar más, o bien laborando menos horas que la jornada legal cumplen con las otras dos condiciones) y subempleados (personas que laboran menos que la jornada legal y desean trabajar más horas, o bien trabajan mínimo la jornada legal pero sus ingresos son menores al salario unificado, entre otras situaciones). A partir del 2007 se agrega la categoría de ocupados no clasificados (aquellos que no entran en las categorías de ocupados plenos u otras formas de subempleo debido a la falta de datos de ingresos u horas laboradas).

27. **Promedio de Años de Escolaridad Terminados de los Ocupados Plenos:** Muestra el promedio de años de escolaridad de la población ocupada plena en un año dado. La población plenamente ocupada es aquella que labora como mínimo la jornada legal, tiene salarios mayores que el salario unificado y no desea trabajar más, o bien laborando menos horas que la jornada legal cumplen con las otras dos condiciones.

28. **Promedio de Años de Escolaridad Terminados de los Subempleados:** Muestra el promedio de años de escolaridad de la población subempleada en un año dado. La población subempleada es aquella que labora menos que la jornada legal y desea trabajar más horas, o bien trabaja mínimo la jornada legal pero sus ingresos son menores al salario unificado, entre otras situaciones).

29. **Promedio de Años de Escolaridad Terminados de los Desempleados:**
Muestra el promedio de años de escolaridad de la población desempleada en un año dado. Desde el 2007 con la nueva metodología aplicada, el desempleo es la suma del desempleo abierto (sin empleo y buscando uno) y el desempleo oculto (sin empleo, no busca uno pero si está disponible para trabajar).