

CAPÍTULO 4 METODOLOGÍA

4.1 Planteamiento de la hipótesis

Como se mostró a partir de la evidencia empírica, la mayoría de los trabajos sobre convergencia aplicados a los estados que conforman la República Mexicana llegan a la conclusión de que antes de 1985 existía una convergencia regional, es decir, antes de la apertura de la economía mexicana (con el ingreso del país al Acuerdo General sobre Aranceles Aduaneros y Comercio, GATT). Después con la apertura de la economía y principalmente después de la firma y entrada en vigor del TLCAN (en 1994) los estados mexicanos presentan un proceso de divergencia real con la acentuación de disparidades regionales (Calderón y Tykhonenko, 2006).

Los trabajos que se mencionan en la evidencia empírica y que llegan a la conclusión antes mencionada, manejan datos hasta el 2001, así que es necesario hacer pruebas de lo que ha sucedido en los últimos años, específicamente desde la aplicación del Tratado de Libre Comercio (TLC). Además todos analizan convergencia absoluta es necesario hacer una análisis de convergencia condicional para tomar en cuenta las diferencias de cada entidad federativa en cuanto a tasas de crecimiento de la población, ahorro, depreciación y cambio tecnológico.

Debido a que ya han pasado varios años desde la apertura comercial en México, esperaríamos observar los efectos del largo plazo mencionados en el modelo de Solow, esto es que haya convergencia. Tomando en cuenta lo anteriores, la hipótesis que vamos a contrastar es que **con la apertura comercial debida a la entrada en vigor del Tratado de Libre Comercio (TLC), el crecimiento de los estados converge.**

4.2 Objetivo general

Para comprobar la hipótesis anterior se emplearan tres metodologías basadas en la teoría de la convergencia. **La primera** es estimar la convergencia absoluta de corte transversal y **la segunda** es estimar la convergencia condicional con panel de datos, que son las formas básicas para el análisis. Y **la tercera** que se usa es la convergencia tipo sigma, indica que

tan dispersos se encuentran los datos de su valor medio en cada año. A continuación se explica con mayor detalle cómo se aplicó cada una.

4.3 Descripción de los modelos y métodos.

En esta sección se explican los modelos y métodos que vamos a usar. Es relevante para esta sección la información sobre panel de datos, ya que no se habían mencionado anteriormente en el trabajo. Para el caso de convergencia absoluta y condicional, en el marco teórico se explica a detalle cómo obtener las ecuaciones, por lo que aquí sólo se muestra la regresión que necesitamos aplicar.

4.3.1 Convergencia absoluta con datos de corte transversal

La regresión para analizar la convergencia absoluta usando corte transversal es la siguiente (Messmacher, 2000):

$$(\ln y_t - \ln y_{t-1}) = \beta_0 + \beta_1 \ln y_{t-1} + u_t \quad (4.1)$$

$(\ln y_t - \ln y_{t-1})$ es la tasa de crecimiento del PIB *per capita* en el periodo de tiempo t_2-t_1 , mientras que y_{t-1} es el nivel inicial de PIB *per capita* y β_0 es el intercepto. Como se explicó anteriormente, el parámetro que tenemos que observar es β_1 , si es negativa indica que hay convergencia, en caso contrario es divergencia.

4.3.2 Convergencia condicional con panel de datos

El uso de panel de datos puede solucionar los problemas del análisis con corte transversal y es otra de las formas que se han usado para el análisis de convergencia. Los cálculos con base en los datos de panel permiten capturar la naturaleza de la dinámica transitorio del modelo de convergencia al utilizar información más rica que combina tanto la dimensión individual, que ofrecen los datos en cortes, como la temporal, que ofrecen las series cronológicas (Calderón y Tykhonenko, 2006), lo que permite introducir el aspecto dinámico del crecimiento y resolver el problema vinculado con la heterogeneidad.

Los enfoques de los datos de panel constituyen una alternativa de los enfoques en corte transversal. En un panel de datos se da seguimiento a la misma unidad transversal

(individuos, empresas, familia, estado, etc.) a través del tiempo. Por lo tanto podemos analizar espacio y tiempo.

Se dice que hay un panel balanceado cuando contamos con los mismos periodos T para cada una de las unidades de corte transversal N. Existen varios métodos para trabajar con un panel de datos, el primero y más simple es aplicar *mínimos cuadrados ordinarios*, sin embargo, éste omite las dimensiones de espacio y tiempo. Otro método que sí aprovecha las características de un panel de datos es el de efectos fijos, se usa cuando se sospecha que existe un efecto asociado a cada unidad transversal, inobservable y constante en el tiempo. Este método nos permite eliminar ese efecto (Mayorga y Muñoz, 2000). Por otro lado, si se sospecha que el efecto inobservable no es constante en el tiempo, se usa efectos aleatorios.

En la evidencia empírica se mencionó el trabajo de Carrillo *et. al.*, (2007), donde encontraron que la educación es un factor importante para determinar la convergencia entre los estados por lo tanto vamos a agregar a la regresión la variable capital humano. De la ecuación 2.42 vamos a agregar al vector x_{it} la variable educación. Este es conocido como el modelo de Solow aumentado (Bond, Hoeffler y Temple, 2001)

4.3.3 Efectos fijos

Para entender cómo funcionan los *efectos fijos* tenemos las siguientes ecuaciones. Consideremos el modelo con una sola variable explicativa para cada i, donde a_i es el factor inobservable constante en el tiempo para cada i.

$$y_{it} = \beta_1 x_{it} + a_i + u_{it}, \quad t = 1, 2, \dots, T \quad (4.2)$$

También para cada i, promediamos la ecuación en el tiempo y tenemos:

$$\bar{y}_i = \beta_1 \bar{x}_i + a_i + \bar{u}_i, \quad (4.3)$$

Donde $\bar{y}_{it} = T^{-1} \sum_{t=1}^T y_{it}$, a_i permanece constante en el tiempo, por eso aparece en ambas ecuaciones. Si sustraemos la ecuación 4.3 de la 4.2 para cada t, obtendríamos, según Wooldrige (2001):

$$y_{it} - \bar{y}_i = \beta_1 (x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i \quad t = 1, 2, \dots, T \quad (4.4)$$

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it} + \ddot{u}_{it}, \quad t = 1, 2, \dots, T \quad (4.5)$$

Donde $\ddot{y}_{it} = y_{it} - \bar{y}_i$ son los datos centrados para “y”, y de igual manera para \ddot{x}_{it} y \ddot{u}_{it} . Como se puede observar, desapareció el efecto inobservable (por ser constante). Tanto para y como para x utilizamos sus promedios en el tiempo y se hace una regresión con los datos de corte transversal, a la que se le llama el *estimador entre grupos*.

Se dice que el estimador de Efectos Fijos permite una correlación arbitraria entre a_i y las variables explicativas en cualquier periodo. (Wooldridge, 2001). Supongamos que la variable explicativa x_{it} y que a_i se correlacionan. Si se ignora esta situación y se aplica MCO, a_i pasaría a formar parte del error u_{it} , por lo que ahora tenemos lo que se llama un error compuesto $v_{it} = a_i + u_{it}$.¹ Así, se pueden obtener los mismos resultados de efectos fijos si se aplican variables ficticias o dicótomas para cada unidad transversal, en este caso para cada estado (Montero, 2007). Este tipo de variable indica la presencia o ausencia de una cualidad o atributo, toma los valores de 0 y 1, donde 1 indicará la presencia de ese atributo y 0 la ausencia de tal atributo; ahora bien, este no es un método recomendado para conjuntos de datos de panel con muchas observaciones de corte transversal.²

Las variables dicótomas obtienen el valor medio de las distintas categorías, por lo tanto, no es una novedad que podamos obtener los mismos coeficientes para las variables explicativas que obtuvimos en la diferenciación. Con las ficticias queremos saber si una cualidad tiene un impacto significativo en la variable dependiente. En este caso, queremos saber si el hecho de ser diferentes estados influye en la determinación del PIB *per capita*, por lo que tendríamos una intercepción diferente para cada uno. La ventaja que ofrece la regresión de variables ficticias, en comparación con la diferenciación, es que podemos calcular el efecto inobservable a_i para cada i , si es necesario obtener estos valores.

Asimismo, hay tres razones que justifican que efectos fijos es el mejor método para la base de datos de este trabajo. La primera implica que es lógico pensar que los estados

¹ Dado los supuestos de Gauss Markov para el análisis de regresión, que v_{it} no se debe correlacionar con x_{it} , sin embargo al aplicar MCO estamos pasando por alto este supuesto, x_{it} se correlaciona con a_i que está en v_{it} , entonces nuestros estimadores serán sesgados e inconsistentes. Es por este motivo que cuando existe correlación entre el efecto fijo inobservable a_i y las variables explicativas x_{it} , es necesario aplicar efectos fijos, para eliminar a_i y obtener buenos estimadores.

² Otro detalle importante es que con este método obtenemos una R cuadrada elevada, pero no debemos emocionarnos, esto se debe a que estamos incluyendo muchas variables ficticias.

tienen características constantes en el tiempo y que los hacen diferentes. Sería muy difícil tomar en cuenta las características de cada uno de los estados, la complicación es mayor si estas características no son observables, por lo tanto, efectos fijos nos ayuda a eliminar estos datos y simplificar el modelo. La segunda razón es que muchos de los trabajos que usan panel de datos para analizar convergencia encuentran que efectos fijos es un método adecuado. La tercera es más formal, pues implica decidir entre aplicar efectos fijos o aleatorio, para lo que se realiza el *test de Hausman* que a continuación se explica.

4.3.4 Test de Hausman

El test propuesto por Hausman (1978) es un test *chi-cuadrado* que determina si las diferencias son sistemáticas y significativas entre dos estimaciones. Se emplea en dos casos (1) para saber si un estimador es consistente, (2) si una variable es o no relevante. Esta prueba se puede realizar con cualesquiera dos modelos de regresión que queramos comparar. (Montero, 2005).

En este caso la prueba de Hausman permite comprobar si debemos usar *efectos fijos o aleatorios*. Como se explicó en la sección anterior cuando existe correlación entre a_i y las variables independientes x , se debe corregir el problema para que los resultados sean insesgados y consistentes. Por lo tanto si existe correlación y no se corrige por *efectos fijos*, los estimadores no van a ser consistentes³.

El test de Hausman compara un modelo de regresión donde se corrige el problema de correlación (*efectos fijos*) y otro modelo donde no se corrige (*efectos aleatorios*). Si entre estos dos no existe una diferencia significativa quiere decir que no hay necesidad de corregir correlación entre a_i y las x , porque no existe tal problema, entonces ambos estimadores son consistentes. Sin embargo, si existe una diferencia significativa entre *efectos fijos* y *aleatorios* quiere decir que si existe el problema de correlación y es importante resolverlo. Debemos aplicar efectos fijos en vez de efectos aleatorios para que el resultado sea consistente. La hipótesis nula de la prueba de Hausman es que los estimadores de efectos aleatorios y de efectos fijos no difieren sustancialmente. Si se rechaza la H_0 , los

³ No van a ser consistentes por que no se cumple un supuesto de Gauss Markov que establece que las variables explicativas x no se deben de correlacionar con el término de error. Al no cumplir con este supuesto los estimadores ya no son MELI (los mejores estimadores lineales insesgados).

estimadores sí difieren, y la conclusión es efectos fijos es más conveniente que efectos aleatorios (Márquez, 2005). Si $p < 0.05$ se rechaza la hipótesis nula de igualdad al 95% de confianza y se deben de asumir las estimaciones de efectos fijos. Si p es mayor que 0.05 no se rechaza la hipótesis nula, entonces ambos estimadores son consistentes pero nos quedamos con *efectos aleatorios*, ya que no utiliza variables dicótomas como el otro método.

La fórmula de la prueba de Hausman es (Wooldridge, 2002,):

$$H = (\hat{\delta}_{FE} - \hat{\delta}_{RE})' [Avar(\hat{\delta}_{FE}) - Avar(\hat{\delta}_{RE})]^{-1} (\hat{\delta}_{FE} - \hat{\delta}_{RE}), \quad H \sim \chi^2_n \quad (4.6)$$

Donde: $\hat{\delta}_{FE}$ es el vector de estimaciones del estimador consistente (efectos fijos); $\hat{\delta}_{RE}$ es el vector de estimaciones del estimador eficiente (efectos aleatorios); $Avar(\hat{\delta}_{FE})$ es la matriz de covarianza del estimador consistente; $Avar(\hat{\delta}_{RE})$ es la matriz de covarianza del estimador eficiente; y n son los grados de libertad de la χ^2_n . La ecuación (4.6) calcula las diferencias en las estimaciones comunes a ambos modelo.

4.4 Problemas inherentes en las regresiones de convergencia condicional

Cualquier investigación empírica deberá enfrentarse a sustanciales problemas tanto en la estimación, como en la interpretación del crecimiento económico, esto son (Carrillo, *et. al.*, 2007):

1. *Heterogeneidad en los parámetros*. Una de las principales críticas al modelo de convergencia absoluta es que los países difieren ampliamente en aspecto culturales, sociales, políticos e institucionales que los llevan a tener valores de sus variables en el largo plazo (en estado estacionario) divergente. La utilización de variables promedio en este caso puede ser positiva, debido a que el propósito del trabajo de corte transversal es llegar a generalizaciones acerca del crecimiento. Las regresiones de este tipo, siguiendo a Temple (1999) se realizan usando una estimación de datos de panel estático, donde la variación de los coeficientes es aleatoria e independiente de los regresores.
2. *Valores extremos (outliers)*. Podrían existir observaciones aberrantes u *outliers*. Sin embargo no es correcto sólo deshacerse de las observaciones sin ningún rigor que conlleve alguna metodología estadística que demuestre que esa

observación no es representativa. Una alternativa es la de la estimación robusta, debido a que ésta permite ubicar a la mayor parte coherente de nuestra serie de datos, y el usar los parámetros estimados para identificar posibles outliers.

3. *Endogeneidad*. Es el problema más preocupante en los trabajos de crecimiento, la endogeneidad de algunos regresores usados como *proxies* de los determinantes del crecimiento. Una forma de solucionarlo ha sido el usar valores iniciales de las variables, otras alternativas son el uso de valores rezagados de las variables o de variables instrumentales, y correr MCO en dos etapas.

