

3. VALIDACIÓN DE HIPÓTESIS

Los anteriores capítulos han ido estableciendo el marco de la investigación objeto de la presente tesis. De forma descriptiva y analítica hemos examinado la evolución demográfica, obteniendo conclusiones o deducciones que han de verse refrendadas ahora por los métodos cuantitativos utilizados en econometría. No obstante, antes de pasar a la aplicación de los mismos conviene destacar algunos aspectos contenidos en los capítulos que anteceden.

En el Apéndice A se efectúa una completa revisión de las teorías de la población, de resultas de la cual se obtiene un panorama disperso en cuanto a los factores explicativos de la dinámica demográfica. Una gran amplitud de causas se baraja por los diversos autores, generalmente en explicaciones monocausales, contradictorias unas con otras las más de las veces y abarcando desde la estricta independencia de leyes generales a las peregrinas dependencias de características mentales cambiantes. Puede extraerse un núcleo de acuerdo relativo a la complejidad de causas que influyen en la dinámica poblacional, cuya actuación conjunta y simultánea sería norma general, y al reconocimiento de factores económicos como causas activas, no únicas, de la evolución demográfica.

En definitiva, la revisión teórica del Apéndice A pone de manifiesto el débil estado de la teoría de la población, con pocos principios asentados o aceptados de forma generalizada. Y el análisis cuantitativo que queremos desarrollar en este capítulo se resiente desde el principio si la base teórica subyacente es débil, porque, como fundamento metodológico, hemos de recordar que, "la teoría debe ser la guía para la construcción de cualquier modelo". Ante el amplio panorama de planteamientos diversos y la controversia suscitada por la mayoría de las teorías de la población, elegimos como objetivo la contrastación de dos postulados muy presentes (a veces, sólo de forma

tácita) en las obras demográficas, históricas, sociológicas o económicas: que la población no es una variable independiente y que ciertas variables económicas influyen significativamente en la población.

El capítulo primero inició el abordaje de esos postulados desde dos puntos de vista: analítico y empírico. En el análisis de las funciones matemáticas como leyes generales de la población concluíamos con la no adecuación de dichas leyes para la descripción de la dinámica demográfica, minando por tanto la propia base de la independencia de la variable población. Un segundo apartado, dedicado a la búsqueda de evidencias del carácter dependiente de la población recogía todo un catálogo de expresiones de apoyo sobre la dependencia, formuladas por un variado plantel de autores de diversas disciplinas, completado con una revisión histórica de la población española que permitía explicar cada uno de los altibajos demográficos en función de causas concretas, poniendo especial énfasis en las de tipo económico.

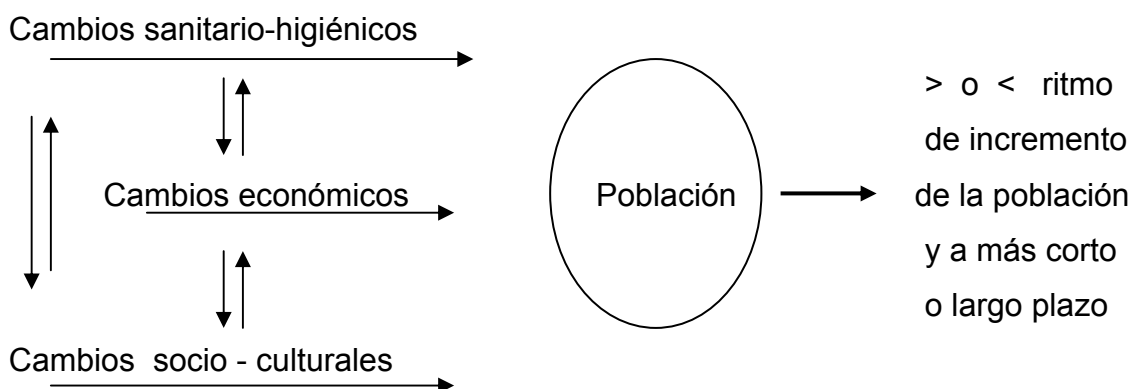
La importancia del capítulo segundo radica en que esclarece el mecanismo de evolución de una población, en base a un modelo, que quiere ser realista, desarrollado en el programa informático "Pirámide". Los estudios llevados a cabo con el modelo nos permitían conocer cómo la población *dosifica* la respuesta ante determinado estímulo en función de su estructura por edades, cómo algunas respuestas se manifestaban con perfil oscilante, cómo graduales cambios no provocaban consecuencias relevantes inmediatas pero adquirirían importancia cuantitativa a largo plazo y cómo la transición demográfica toma carta de naturaleza como evolución general al desaparecer la mortalidad catastrófica.

Con los conocimientos adquiridos en los anteriores capítulos podemos disponernos a reforzar nuestras observaciones mediante el empleo de métodos cuantitativos, en especial el análisis de regresión. Pero antes de comenzar con el manejo de series de datos y técnicas econométricas, conviene una reflexión sobre el planteamiento subyacente en este capítulo.

3.1 PLANTEAMIENTOS PREVIOS

3.1.1 Naturaleza del problema

Puesto que mantenemos que una población no evoluciona en el tiempo de forma independiente a las circunstancias socio-económicas a las que ha de enfrentarse, podemos elaborar un esquema basado en el mecanismo causa-efecto para estudiar inicialmente su aplicación al caso del desenvolvimiento de la población. En forma general el esquema sería el siguiente:



Una serie de cambios, que hemos resumido en tres tipos generales, incidirían sobre la población, serían las causas que, a modo de shocks, provocarían efectos en la variable objeto de estudio, que no es otra que el nivel de población. Aunque los efectos pueden instrumentarse de diversa formas, tales como variación en la mortalidad, en la fecundidad, emigración, etc., se manifiestan en última instancia como aumento o disminución del ritmo de incremento de población, manifestación que tendrá lugar a corto, medio o largo plazo, circunstancia que forma parte del propio efecto asociado a las causas que actuaron.

Aunque nuestra preocupación posterior estará dirigida a las causas económicas, ampliemos de momento la visión considerando todo el conjunto de cambios, causas o shocks que se recogen en el esquema. Pueden observarse las dobles flechas entre cada grupo de cambios, para poner de

manifiesto la *interrelación* entre ellas: no sólo actúan de forma conjunta, si no que determinados cambios sanitario-higiénicos tendrán repercusiones económicas (a modo de ejemplo, una mejora sanitaria, al reducir las enfermedades, aumentará la productividad), lo mismo que ciertos cambios económicos producirán modificaciones socio-culturales (a modo de ejemplo, un aumento de la renta per cápita inducirá un mayor consumo e inclinación por el ocio). Por tanto un cambio económico, o sanitario, o socio-cultural, no sólo actuará sobre la población por su impacto directo, sino que de forma indirecta, ampliando o reduciendo los impactos de los otros grupos, afectará también al nivel de población.

Esta circunstancia de interrelación entre todo tipo de causas establece una inicial dificultad para un análisis cuantitativo que requiere manejar datos individualizados, por cuanto es muy difícil separar el impacto directo debido a un tipo de causas del indirecto ocasionado a través de los restantes tipos de cambios o causas. Como nuestra preocupación es puramente económica tendremos que englobar el resto de factores causales en el término estocástico de perturbación, V_t , adulterando desde el principio un supuesto básico en el modelo de regresión lineal como es el que señala que la perturbación, V_t , y la(s) variable(s) explicativa(s) no están correlacionadas. Esto nos obligará a utilizar, en lugar del supuesto fuerte de que las X_t y las V_t son independientes, un supuesto más débil de que los valores de las variables X_t y V_t no están correlacionadas contemporáneamente (es decir, en el mismo momento del tiempo). En este caso los estimadores MCO pueden estar sesgados pero son consistentes. Y este supuesto débil se cumplirá en nuestro planteamiento por cuanto los impactos indirectos a través de otros grupos de cambios no coincidirán en el tiempo con los impactos directos de los cambios económicos. En cualquier caso en la práctica econométrica suele prescindirse de esta vulneración del supuesto de independencia entre las variables explicativas y las perturbaciones, invocando tan sólo el carácter no estocástico de las variables explicativas, cuyos valores están dados.

La complejidad del problema aumenta al considerar la forma en que la población responde ante esos cambios o causas económicas. El impacto no

siempre tendrá reflejo cuantitativo, porque puede ser absorbido o tener respuesta cualitativa (ante un aumento de renta el nivel de fecundidad puede permanecer impasible, o sufrir un descenso compensado por un proporcional descenso de la mortalidad, o provocar que unas familias adelanten nacimientos y otras, en igual medida, los retrasen) y en esos casos no tendremos percepción numérica utilizable. Si la variable observable es el nivel de población no podremos percibir los efectos consistentes en cambios en la estructura de edades, retrasos en la maternidad, variaciones en la fecundidad y en la mortalidad, sino cuando tengan posterior actuación acelerando o frenando el ritmo de incremento de la población. Debemos deducir, pues, la necesidad de incorporar retardos entre series de datos temporales relativos a causas económicas y las correspondientes series de la variable dependiente población.

El problema a resolver es aclarar la influencia de ciertas variables económicas sobre otras demográficas, reconociendo la actuación conjunta de múltiples causas, muchas de ellas no estrictamente económicas. Este relativismo realista también reconoce la compensación de efectos derivados de causas individuales, lo cual dificulta la investigación, pensando, no obstante, que es posible establecer mecanismos causales según el esquema conocido que recoge que los cambios en las variables exógenas (económicas) se comportan como "shocks" que causan los cambios en la variable endógena (población). Y todo ello con el gran obstáculo ya citado de que el esquema estímulo-respuesta no es fácilmente observable porque la población puede absorber el estímulo, retrasar la respuesta o ésta puede producirse en términos no cuantitativos, o bien las respuestas se equilibran (menor mortalidad y menor fecundidad).

Conocida la magnitud del problema y considerando que los métodos cuantitativos son procedimientos para "interrogar" a los datos disponibles para que den información, hay que pensar que en este caso necesitaremos ser muy persuasivos y constantes para obtener una declaración medianamente satisfactoria, pero de forma "legal": el objetivo es probar las hipótesis expuestas a priori en lugar de imponerlas simplemente sobre los datos para los cuales ellas pueden no ser ciertas.

3.1.2 Características y fiabilidad de los datos

En econometría existe conflicto entre la necesidad de disponer de amplias series de datos y la creciente falta de homogeneidad al aumentar el tamaño de la serie (temporal). Los métodos cuantitativos proporcionan resultados más robustos si contamos con un número de observaciones muestrales por encima de 30, a veces más, disponibilidad que puede ser posible en los estudios de corte transversal, pero algo más difícil de lograr en el trabajo con series temporales. Y cuando, en casos excepcionales, podemos disponer de largas series de tiempo, aparece la indeseada compañía de falta de homogeneidad, tanto mayor cuanto más atrás en el tiempo se remonte la serie de datos.

La falta de homogeneidad se debe a diversas causas: variación en las definiciones de magnitudes a cuantificar, en los criterios de estimación u observación, en los instrumentos analíticos de cálculo o ponderación o en la propia extensión de muestreo o recuento del suceso bajo estudio. La evolución de los métodos estadísticos y las mejoras en las aplicaciones de los mismos han logrado en los últimos años ganancias de rigor que pueden llevar a que sean poco comparables las mediciones de la misma magnitud en momentos temporales distintos, en definitiva a que la serie "bruta" emitida por un mismo organismo presente falta de homogeneidad.

Para compensar las variaciones en los métodos o criterios de obtención de datos se les suele someter, posteriormente, a una operación de homogeneización o corrección, tratando de presentar una serie final de datos homogéneos. No siempre es explícita esta operación de homogeneizar la serie, ni se aplica en igual forma por distintos autores u organismos estadísticos, lo cual explica las variaciones encontradas en diferentes textos y publicaciones al referirse a series de la misma magnitud y período temporal. En todo caso la situación sólo adquiere gravedad cuando la serie abarca períodos históricos conflictivos (guerra civil española: la excepcionalidad y carencia de datos requieren de complementos de gran carga subjetiva) o muy alejados del

presente, cuando aún no estaban afianzados ni extendidos los procedimientos estadísticos, en particular, en España, antes de 1950.

Poner de manifiesto estas dificultades no ha de impedirnos la utilización de algunas series que recorren todo el siglo XX español, si bien quedan apuntadas las reservas en cuanto al grado de homogeneidad que pueden presentar. Haremos, en todo caso, mención de los procesos de homogeneización a que han sido sometidos los datos y reservaremos un análisis más exhaustivo para series más modernas, libres del pecado de heterogeneidad. Como ocurre a todo científico social dedicado a una labor de análisis de datos (obtenidos por otros investigadores, analistas u organismos estadísticos) hemos de utilizar lo que tenemos, porque en el reparto del trabajo unos se dedican a la medición u obtención de datos y los econométricos al manejo posterior de los mismos. Y aunque no está justificada la excusa de achacar a los "malos datos" la insuficiencia explicativa de nuestros modelos o resultados, ocurre, sencillamente, que a veces nuestra materia prima deja mucho que desear.

Otro problema bastante común en las aplicaciones prácticas econométricas es el que aparece ligado a la utilización de series de tiempo, en particular económicas: dada la normal inercia de los acontecimientos económicos, que hace que se mantengan los efectos de las acciones pasadas sobre situaciones actuales y futuras, aparece la autocorrelación o correlación referida a los valores de una misma serie, incumpléndose una de las hipótesis del modelo clásico de regresión lineal. Y el efecto principal es la invalidez del método de estimación por mínimos cuadrados ordinarios (MCO) por cuanto los estimadores obtenidos, si bien siguen disfrutando de las propiedades de linealidad, insesgadez y consistencia, ya no son eficientes, siendo de dudosa aplicación los test y contrastes estadísticos.

Merece la pena una mayor insistencia en la posibilidad de que las perturbaciones de sucesivos períodos estén correlacionadas porque nuestro posterior análisis cuantitativo utilizará principalmente series temporales. Para explicar la inercia de los acontecimientos económicos podemos pensar en la

aparición de una nueva tecnología en el momento t , que nos produce un efecto no esperado en dicho momento, pero, además, dicho efecto no desaparecerá en $t+1$, lo cual implica autocorrelación o correlación entre las perturbaciones de ambos momentos. Similares consecuencias aparecen asociadas a la presencia de tendencias: la incorporación laboral de la mujer se manifiesta siguiendo un patrón definido, con valores relacionados entre momentos de tiempo consecutivos.

Las series de tiempo, material básico de nuestro análisis, aún arrastran otra complicación si, como sucede en la práctica, no son estacionarias. La mayor parte del trabajo empírico basado en datos de series temporales supone que éstas son estacionarias, por lo que un primer presupuesto al manejar series temporales es el de revisar su estacionariedad. Precisamente para corregir la no estacionariedad se somete la serie original a métodos de alisado, transformaciones que tienden a suavizar los datos, cálculo de diferencias, etc., procesos que pueden crear esquemas de autocorrelación que se englobarán en la perturbación aleatoria. De una u otra manera las series económicas temporales llevan aparejada la autocorrelación, como tendremos ocasión de comprobar a través de diversos contrastes, obligándonos a la utilización del método de mínimos cuadrados generalizados (MCG) para la estimación de los parámetros del modelo.

Puesta de manifiesto la presencia o "compañía" de autocorrelación en las series económicas temporales, ¿ocurre otro tanto con las series de población? Fundamentalmente, sí, porque lo que hemos denominado normal inercia de los acontecimientos económicos tiene su plasmación en el campo demográfico como tuvimos ocasión de observar en el capítulo cuarto, al recoger las consecuencias oscilantes o cíclicas, que se extendían en el tiempo como respuesta a un impacto en un único momento t . Y además, tenemos el añadido problema de la interpolación. Una fuente de autocorrelación puede ser la interpolación: sabemos que los Censos de Población se elaboran cada diez años, ahora bien, si se necesita obtener datos para algún año comprendido en el período intercensal 1950-1960, por ejemplo, la práctica común consiste en la interpolación con base en algunos supuestos ad hoc. Esta técnica de "manejo"

podría imponer sobre la información un patrón sistemático que pudiera no estar presente en la información original.

Dada la necesidad de contar con series anuales de población, pese a contar sólo con datos decenales de directa extracción, se obtienen los valores de años intercensales por interpolación. Así han sido elaboradas las series de población que manejaremos, de forma que tendremos que convivir con la autocorrelación como pesado personaje no invitado, recurriendo, cuando sea necesario, a la estimación por otros métodos (MCG).

Nuestra serie básica referida a la población abarcará desde 1900 a 1996, tomada de la obra de Angus Maddison "La economía mundial 1820-1992" editada por Perspectivas OCDE, que recoge los datos decenales y las interpolaciones llevadas a cabo por la OCDE. Los últimos dos años de la serie han sido añadidos de datos tomados del INE. Del total de 97 datos de la serie, sólo diez serían datos primarios, los correspondientes a los censos decenales, a los que podemos añadir las tres observaciones relativas a los padrones de 1975, 1986 y 1996. El resto, un total de 84 datos, no son de directa observación, sino de deducción a partir de los datos primarios.

¿Podemos pensar que, al menos esos datos primarios referidos a los censos, son totalmente fiables? Hay que responder que los datos no son igualmente fiables en todas las épocas. Amado de Miguel describe en su obra "España cíclica", pág. 18, un índice de bondad relativa al censo en cuestión que "consiste en poner en relación las *personas censadas menores de un año* con el *número de nacidos en el año menos los fallecidos de menos de un año*". Los valores que así obtiene son:

Año	Índice de bondad	Año	Índice de bondad
1900	77	1950	86
1910	81	1960	93
1920	85	1970	101
1930	87	1980	107
1940	78		

Podemos apreciar que cuanto más atrás en el tiempo los censos infraestiman notoriamente más el número de niños pequeños, de donde puede deducirse que de las posibles desinformaciones de los censos, una muy característica debe ser la de no dar cuenta de los niños pequeños. En los últimos censos aumenta decididamente la "bondad", de forma que a partir de 1960 puede admitirse que son ya razonablemente fiables. La sobre-bondad del censo de 1981 resulta inexplicable, debiendo atribuirse en mayor medida a un error del numerador porque se consideran más fiables las inscripciones en el Registro Civil (nacidos y fallecidos del denominador) que las anotaciones de las hojas censales.

El censo de 1940 presenta un retroceso en la creciente fiabilidad de los censos, tendencia interrumpida con la Guerra Civil. Está generalmente aceptada la mayor relatividad de las cifras de este censo concreto, por las mayores dificultades de elaboración tras el período bélico. Encontramos a este respecto en Historia de España de editorial Planeta, tomo XII, página 78: " El censo de población de 1940, netamente superior al de 1930 desborda ampliamente la realidad: la inclusión de personas fallecidas o exiliadas, las dobles inscripciones -en la mayoría de los casos debidas a deficiencias técnicas- incrementaron en varios cientos de miles la estadística oficial".

Como serie económica básica manejaremos la serie del PIB desde 1900 a 1996, extraída asimismo de la obra de Angus Maddison "La economía mundial 1820-1992", editada por Perspectivas OCDE. Se trata de una serie temporal, de datos anuales, valorados en dólares de 1990 y según el enfoque Geary-Khamis para permitir comparaciones multilaterales (está basado en los conceptos gemelos de paridad del poder adquisitivo y precios promedios internacionales de mercancías). El propio autor, Angus Maddison, refleja que la primera parte de la serie, proporcionada por L. Prado de la Escosura, se obtuvo como una estimación gruesa de la variación del ingreso real, mientras que a partir de 1950 se hizo un ajuste fino de los saltos en el nivel del PIB entre segmentos sucesivos de las cuentas nacionales oficiales.

La utilización de esta serie concreta responde al respaldo que le confiere la aceptación de un organismo internacional, como es la OCDE, para comparaciones internacionales, habiendo sido deflactada de acuerdo con métodos de general aplicación en dicho organismo. Nuevamente hay que resaltar que el período correspondiente a la Guerra Civil, e incluso el de inmediata posguerra, es sólo relativamente fiable, y ello por las propias circunstancias bélicas y la inexistencia de una auténtica contabilidad nacional. No faltan autores que prescinden, en aras del rigor, de las mediciones referidas a estos años, interrumpiendo la serie entre 1935 y 1942, pero dado nuestro objetivo de análisis cuantitativo necesitamos la continuidad de la serie, siquiera para un estudio general, aceptando, pues, la serie propuesta por Angus Maddison y que sirve de referencia para la OCDE.

El resto de series temporales que emplearemos, referidas a variables económicas y demográficas, arrancan en 1964 y no presentan problemas de fiabilidad (aunque subsiste la cuestión de la autocorrelación). Se emplearán para un estudio más afinado y elaboración de modelos de regresión entre variables demográficas y económicas, con el objetivo de validar la hipótesis número 2, que afirma que "determinadas variables económicas influyen significativamente en la dinámica demográfica". Son series más cortas en el tiempo, pero de aceptada validez, obtenidas de forma homogénea por el mismo organismo, INE, salvo cuando se señale otra procedencia.

El señalamiento efectuado en este apartado respecto a las limitaciones de los datos disponibles en cuanto a problemas de heterogeneidad, presencia de autocorrelación, no estacionariedad y falta de fiabilidad, quiere dejar claro desde el principio que los resultados que puedan obtenerse no dejarán de ser modestos y cuestionables, pese a la correcta aplicación de métodos cuantitativos, objetivo este último de explícita persecución por sí mismo como corresponde al carácter científico de este estudio. Tampoco conviene desdeñar algunos tímidos pasos, por insignificantes que estos sean, si colaboran a ilustrar algunos aspectos y propician el avance científico.

3.2 VALIDANDO LA HIPOTESIS N° 1

El punto de arranque del presente trabajo era la frase final del libro de J. Nadal "La población española. Siglos XVI a XX", que concretábamos como:

HIPOTESIS N° 1: La población española no es una variable independiente y que hemos tratado de establecer durante todo el primer capítulo en un análisis teórico y descriptivo, exponiendo argumentos que corroboraban la no independencia de la población, tanto por resaltar la inadecuación de leyes generales (cuya validez asentaría la independencia), como por una recopilación de observaciones que mostraban la dependencia de la población en relación a otras variables socio-económicas. Apoyábamos allí la no independencia con una colección de opiniones de reputados autores y recogíamos las evidencias de la propia historia española en una revisión de la evolución de la población desde 1860 a nuestros días, en la cual podíamos "explicar" cada secuencia demográfica por la actuación de alguna(s) causa(s) de orden económico, social, político o catastrófico.

Todo lo allí expuesto nos llevaba a la evidencia contenida en la hipótesis n° 1 : La población no es una variable independiente; evidencia que trataremos de complementar en este apartado con el uso de métodos cuantitativos, siguiendo un básico criterio econométrico que afirma que la teoría debe ser la guía para la construcción de modelos. Hemos llegado a conclusiones teóricas que echaban por tierra la pretendida independencia de la población y el trabajo subsiguiente ha de ser corroborar mediante análisis cuantitativo esas conclusiones teóricas.

Dado el carácter general de la negación contenida en la hipótesis n° 1, su contrastación requiere de series que abarquen un largo período de tiempo, para no incurrir en el estudio de cortos períodos para los cuales, de forma extraordinaria, pudiera ser válida la negación por las especiales circunstancias de unos años determinados. Manejaremos, pues, la serie anual de la población española 1900-1996 y la serie anual del PIB correspondiente a esos mismos años, en dólares constantes de 1990. A los inconvenientes reconocidos en el anterior apartado hay que añadir la falta de estabilidad estructural (cuando

menos, se acepta un cambio estructural en torno a 1960) de dichas series, pero con todas las limitaciones conocidas es preferible extendernos en el tiempo si deseamos un análisis general. A continuación se muestran ambas series:

Año	Población	PIB	Año	Población	PIB
1900	18.566	37.873	1949	27651	65.239
1901	18.659	40.112	1950	27.868	66.792
1902	18.788	38.513	1951	28.086	73.874
1903	18.919	39.427	1952	28.332	79.676
1904	19.050	38.604	1953	28.571	80.589
1905	19.133	38.330	1954	28.812	85.204
1906	19.316	40.432	1955	29.056	89.635
1907	19.450	40.843	1956	29.355	96.077
1908	19.585	41.894	1957	29.657	100.188
1909	19.721	42.899	1958	29.962	104.666
1910	19.858	41.620	1959	30.271	102.701
1911	19.994	44.452	1960	30.583	105.123
1912	20.128	43.812	1961	30.904	117.549
1913	20.263	45.686	1962	31.158	128.514
1914	20.398	43.950	1963	31.430	139.752
1915	20.535	44.543	1964	31.741	148.387
1916	20.673	46.234	1965	32.085	162.823
1917	20.811	45.457	1966	32.453	179.727
1918	20.950	45.411	1967	32.850	191.468
1919	21.091	46.097	1968	33.240	208.144
1920	21.232	49.021	1969	33.566	231.535
1921	21.411	50.848	1970	33.876	246.976
1922	21.628	52.538	1971	34.190	259.814
1923	21.847	53.955	1972	34.498	281.560
1924	22.069	56.330	1973	34.810	304.220
1925	22.292	59.894	1974	35.147	321.809
1926	22.518	59.848	1975	35.515	325.007
1927	22.747	64.736	1976	35.937	335.698
1928	22.977	64.508	1977	36.367	346.571
1929	23.210	68.391	1978	36.778	352.190
1930	23.445	65.696	1979	37.108	352.099
1931	23.675	64.234	1980	37.386	356.622
1932	23.897	66.244	1981	37.751	355.754
1933	24.122	65.011	1982	37.961	360.094
1934	24.349	68.117	1983	38.180	366.581
1935	24.579	68.620	1984	38.342	373.160
1936	24.810	54.868	1985	38.505	381.794
1937	25.043	51.168	1986	38.669	394.312
1938	25.279	51.122	1987	38.716	415.556
1939	25.517	54.274	1988	38.809	436.937
1940	25.757	58.934	1989	38.888	456.993
1941	25.979	58.934	1990	38.959	474.125
1942	26.182	62.269	1991	39.025	484.313
1943	26.387	63.868	1992	39.117	488.470
1944	26.594	66.701	1993	39.193	483.582
1945	26.802	61.995	1994	39.308	491.805
1946	27.019	65.422	1995	39.442	505.576
1947	27.223	66.335	1996	39.643	517.204
1948	27.437	65.011			

Fuente: Angus Maddison. La economía mundial 1820-1992

3.2.1 La no independencia estadística

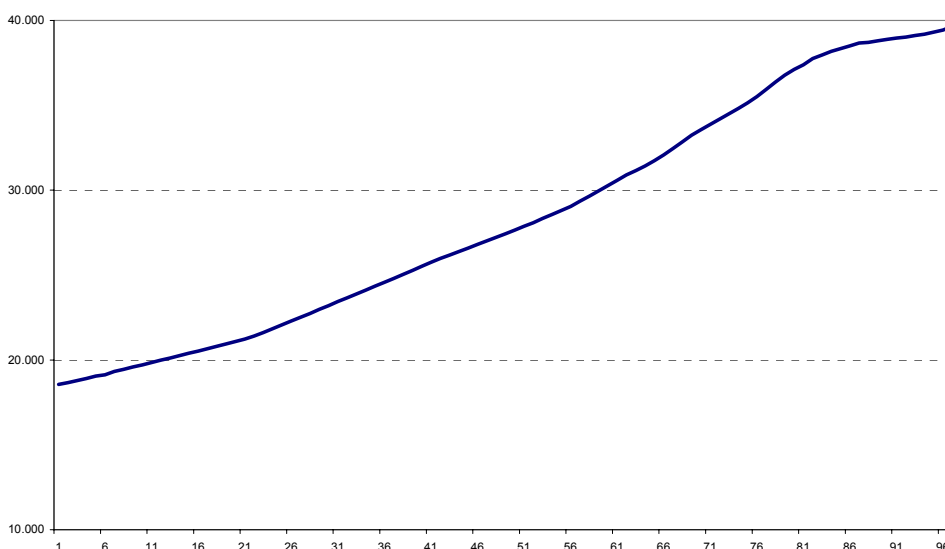
Con las series anteriormente presentadas, representantes de la variable población y la variable economía, queremos demostrar la asociación entre ambas de forma tal que quede descartada la independencia. Para ello nos basamos en una práctica definición de independencia expuesta por Gujarati en su obra "Econometría (1997), página 100:

Para dos variables normalmente distribuidas, una covarianza o correlación cero significa independencia entre las dos variables

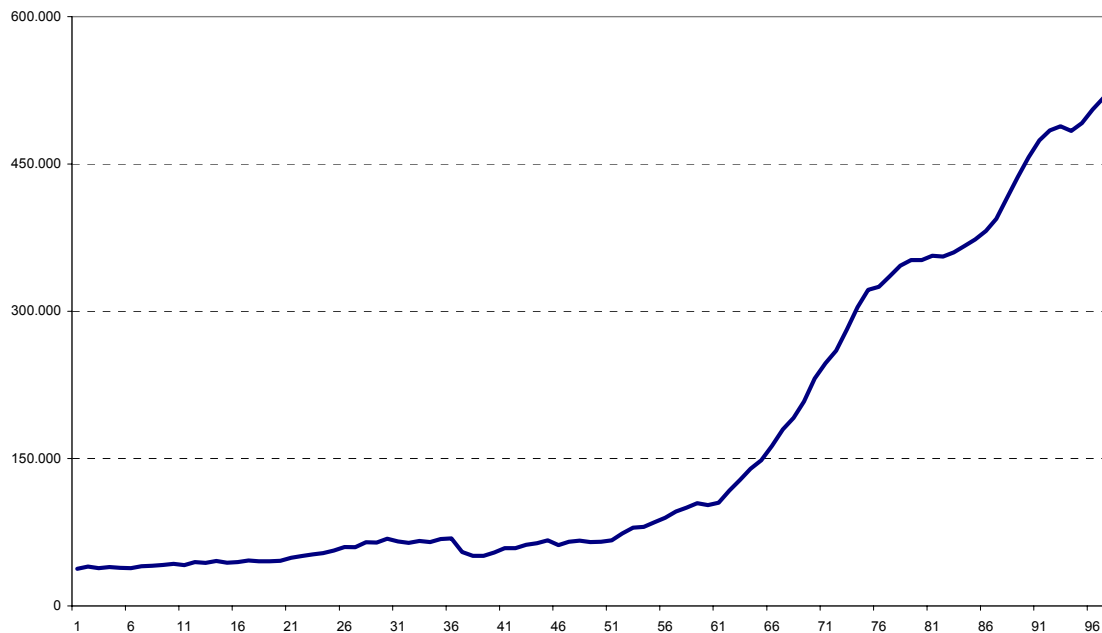
Trataremos de demostrar el incumplimiento de la condición de correlación cero entre nuestras dos series y, por tanto, la falta de independencia entre estas dos variables (población y PIB). La utilización de la serie del PIB como estado o marcha de la economía cuenta con todos los reparos que suelen hacerse a este respecto: es un indicador simple, burdo y limitado, pero pese a todo, no tenemos uno mejor, razón de peso para continuar usándolo, soportando resignadamente sus defectos.

Otra cosa es el manejo analítico de las series. La representación gráfica que ofrecemos a continuación muestra en ambas líneas la presencia de tendencia creciente, una propensión hacia arriba prácticamente constante:

Población española 1900-1996



Evolución del PIB 1900-1996



Nuestras series se mueven en la misma dirección, existiendo entre ambas una propensión creciente, circunstancia frecuente en información relacionada con series de tiempo económicas, y que ha de alertarnos para evitar el problema de **correlación espuria**: cuando dos series tienden a moverse en la misma dirección, reflejando una propensión creciente o decreciente, la regresión efectuada entre ellas producirá un R^2 elevado, valor que puede no reflejar la verdadera asociación entre ellas sino, simplemente, la inclinación común presente en las dos series.

En nuestro caso el diagnóstico de correlación espuria parece de apropiada aplicación, invalidando el elevado valor de R^2 que obtenemos en la regresión entre las dos series originales: $R^2 = 0,838$. La existencia de correlación espuria tiene que ver con el tema mencionado en el apartado anterior, cuando hacíamos referencia a que el análisis empírico basado en información de series de tiempo supone implícitamente que la serie de tiempo bajo análisis es estacionaria. La eliminación de tendencias es uno de los procedimientos utilizado para convertir una serie de tiempo en estacionaria, y por lo tanto, para evitar el problema de la correlación espuria. Seguiremos el método consistente en determinar la tendencia de cada una de nuestras series

mediante regresiones cuyos residuos serán las series libres de tendencia, estacionarias y alejadas del peligro de correlación espuria entre ellas.

Otra cuestión surge al considerar qué tipo de tendencia se ajustará mejor a nuestros datos. Desechando la exponencial y la logística por las propias razones aducidas en el capítulo segundo para mostrar la no adecuación de dichas funciones, nos quedaría elegir entre las polinómicas desde la de primer grado (lineal), segundo grado (cuadrática), tercer grado (cúbica), etc. Entre ellas la función más apropiada para el ajuste analítico de series por el método de los mínimos cuadrados, como pretendemos, debe elegirse sobre la base de ciertos supuestos lógicos relativos a las leyes del proceso de crecimiento, aunque también existen métodos estadísticos que pueden ayudarnos a determinar la forma de la línea de tendencia. A continuación se describe el método expuesto por el economista polaco Oskar Lange en su obra "Introducción a la econometría" (1978), pág. 41 y 42:

Si, por ejemplo, queremos verificar si en el caso dado una línea recta es idónea para ajustar a la serie, primero examinaremos los incrementos de los términos de la serie dada (1ª diferencia). Si dichos incrementos son más o menos iguales podemos tomar una línea recta para el ajuste. Si, sin embargo, los primeros incrementos de la serie no son más o menos los mismos estudiaremos los segundos incrementos (2ª diferencia). En caso de constancia aproximada de los mismos, elegiremos una parábola de 2º grado como línea de tendencia, ya que esa función cumple la propiedad de que los primeros incrementos no son iguales pero son constantes los segundos incrementos. De igual forma se relacionan los terceros incrementos con el polinomio de tercer grado y así sucesivamente.

Aplicando este método a nuestras series encontramos más apropiadas las tendencias cuadráticas o cúbicas en ambos casos, al ser las que muestran mayor constancia de los incrementos asociados a su grado. Nos decidimos por la forma cúbica por estar su media más centrada y ajustarse mejor a procesos de crecimiento como los de población y producción económica que conocen retrocesos en algunos momentos a lo largo del período.

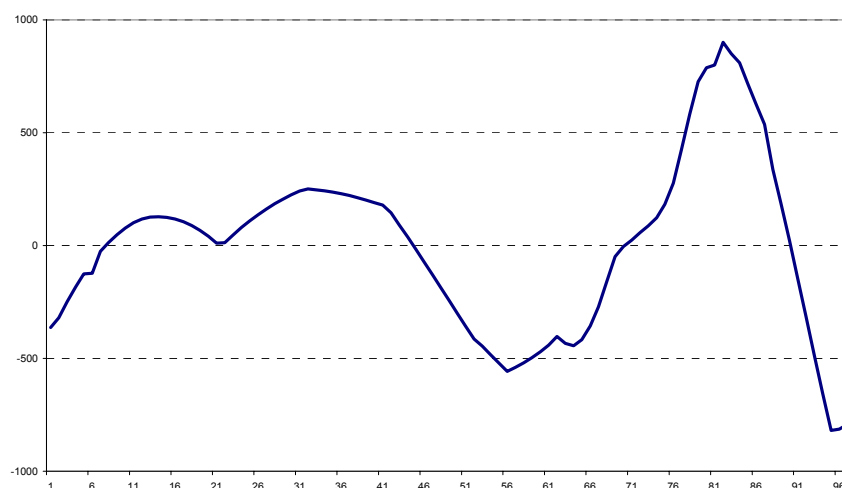
Aceptada la tendencia como función cúbica, en ambas series, procedemos a limpiar las series originales realizando las regresiones siguientes:

$$\text{Pobla} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

$$\text{PIB} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

Los residuos de ambas regresiones constituyen nuestras series limpias de tendencias, que en adelante designaremos por Pobla(I) y PIB(I). En la representación gráfica de las mismas podemos contemplar la persistencia de patrones bien definidos (ciclos), confirmando los supuestos de la metodología de descomposición de los compuestos de una serie, aspecto que, siendo importante, no es objeto del presente estudio y que no afecta a los resultados generales que queremos alcanzar en este apartado.

Serie pobla libre de tendencia



Serie del PIB libre de tendencia



Una vez obtenidas nuestras dos series libres de tendencia, pobla(I) y PIB(I), volvemos a fijarnos en nuestro objetivo: demostrar la no independencia de ambas series basándonos en que cada una presenta una distribución normal y que la correlación conjunta es significativamente distinta de cero, es decir, utilizando el criterio de independencia de Gujarati, contrastar que cumplen individualmente la condición de normalidad, incumpliendo la premisa de correlación cero.

Como prueba práctica de la normalidad de una serie pueden calcularse su asimetría y kurtosis, las cuales deben medir cero y tres, respectivamente, para asemejarse a la distribución normal. Cuanto mayores discrepancias existan respecto a esos valores, menor será la normalidad presente en la serie. Algo más afinado es el test de normalidad de Jarque-Bera que contrasta la hipótesis nula H_0 : "la distribución es normal" mediante el estadístico JB que (asintóticamente) sigue una distribución chi-cuadrado con dos grados de libertad. Dado que las tablas correspondientes muestran el valor 5,99 para un nivel de confianza del 95%, puede concluirse que para valores de JB inferiores a 6 no se rechaza el supuesto de normalidad. Aplicando estas pruebas a nuestras series obtenemos:

Serie: Pobra(I) Población 1900-1996 libre de tendencia
Simetría = 0,149 Kurtosis = 2,942 Contraste JB = 0,374

Serie: PIB(I) PIB 1900-1996 libre de tendencia
Simetría = 0,441 Kurtosis = 2,972 Contraste JB = 3,149

Los valores de asimetría y kurtosis ya permiten deducir una gran afinidad con la distribución normal por su cercanía a cero y tres, respectivamente, en las dos series objeto de estudio. Para mayor conformidad atendemos al contraste de normalidad de Jarque-Bera cuyos valores, 0,374 y 3,149 permiten aceptar la hipótesis nula de normalidad con un grado de confianza superior al 99% en ambos casos. En definitiva, podemos aceptar que nuestras dos series pueden seguir la distribución normal.

El siguiente paso consiste en el cálculo de la covarianza o de la correlación entre las dos series para confirmar su valor distinto de cero. Mediante la conocida fórmula del coeficiente de correlación $r = S_{xy} / S_x \cdot S_y$ obtenemos: $r = 0,715$ valor que nos permite alegar que no existe independencia entre ambas series por su grado de alejamiento del valor nulo (recordemos que el valor de r puede oscilar entre -1 y +1). En consecuencia puede aceptarse la validez de nuestra hipótesis nº 1 que establecía la no independencia de la variable población.

3.2.2 Utilizando el contraste de causalidad

Un punto de vista alternativo puede lograrse mediante la utilización de pruebas de causalidad entre nuestras dos series básicas, no sin antes advertir sobre la intangibilidad de la causalidad para los métodos del análisis cuantitativo. Como indica Gujarati en su obra "Econometría" (1997), pág. 20: "A pesar de que el análisis de la regresión tiene que ver con la dependencia de una variable de otras variables, esto no implica causalidad necesariamente". Continúa citando la obra de Kendall y Stuart "The advanced Theory of Statistics" (1961), pág. 279: "Una relación estadística, sin importar qué tan fuerte y sugestiva sea, nunca podrá establecer una conexión causal; nuestras ideas de causalidad deben venir de estadísticas externas y, en último término, de una u otra teoría".

La conclusión de Gujarati: "una relación estadística no puede por sí misma implicar en forma lógica una causalidad. Para aducir causalidad se debe acudir a consideraciones a priori o teóricas", coincide con lo expuesto por A. Alcaide y N. Alvarez en "Econometría. Modelos deterministas y estocásticos" (1992), pág. 237: "en rigor la causalidad no es contrastable", aclarando que, no se puede, sin más, identificar las variables dependiente e independiente con los conceptos filosóficos de efecto y causa, respectivamente; que la clasificación de variables de acuerdo a ese esquema forma parte de las hipótesis mantenidas o aceptadas.

Todas estas advertencias quieren poner en su justo término la prueba o contraste de causalidad de Granger que aplicaremos a continuación con nuestras dos series de población y PIB libres de tendencia. En todos los manuales econométricos se califica al contraste de Granger como de "causalidad" en cuanto que, estadísticamente, determinaría la dirección de la causalidad entre dos variables entre las que hay una relación temporal. Algún econométra (Edward Leamer) prefiere el término precedencia sobre el de causalidad, precisión que nos parece acertada, aunque siguiendo el uso más extendido seguiremos describiendo la prueba de Granger como "test de causalidad".

Analíticamente, pues, el objeto del contraste de Granger es tratar de determinar el sentido de la causación en el tiempo entre dos variables X e Y, para lo cual se estiman los coeficientes de dos ecuaciones de regresión:

$$Y_t = a \cdot X_{t-1} + b \cdot Y_{t-j} + V_1$$

$$X_t = c \cdot Y_{t-1} + d \cdot X_{t-j} + V_2$$

Donde $j = 1, 2, \dots, n$ representa el número de rezagos o retardos de la propia variable endógena que deseamos incluir en el modelo. El resultado del contraste permitirá establecer la causalidad de Y hacia X si el, o los coeficientes a son significativos (estadísticamente diferentes de cero) sin serlo los coeficientes b. Por el contrario existirá causalidad de X hacia Y si son significativos los coeficientes c sin serlo los d, y hablaremos de retroalimentación o causalidad bilateral cuando los conjuntos de coeficientes de Y y X son estadísticamente significativos, diferentes de cero, en ambas regresiones. Finalmente, se sugiere **independencia** cuando los conjuntos de coeficientes de Y y X no son estadísticamente significativos en ambas regresiones.

Aplicando a nuestras dos series, pobla(I) y PIB(I) la prueba o contraste de Granger, que consiste en el cálculo de un estadístico que sigue la distribución F de Snedecor con m y n-k grados de libertad (m es el número de rezagos, n el número de términos de la serie y k el número de parámetros estimados), obtenemos los siguientes resultados:

Nº de rezagos	Hipótesis nula	F	P
1	Pobla(I) no es causada por PIB(I)	18,777	0,0000
	PIB(I) " " " por Pobla(I)	1,328	0,2521
2	Pobla(I) no es causada por PIB(I)	4,478	0,0140
	PIB(I) " " " por Pobla(I)	2,423	0,0944
3	Pobla(I) no es causada por PIB(I)	6,797	0,0004
	PIB(I) " " " por Pobla(I)	1,925	0,1314
4	Pobla(I) no es causada por PIB(I)	5,255	0,0008
	PIB(I) " " " por Pobla(I)	2,241	0,0714
5	Pobla(I) no es causada por PIB(I)	4,881	0,0006
	PIB(I) " " " por Pobla(I)	2,755	0,0238

En la columna encabezada F se muestran los valores de dicho estadístico obtenidos para cada caso, debiéndose rechazar la hipótesis nula si el valor calculado excede al valor F crítico al nivel seleccionado de significancia, generalmente para un nivel de confianza del 95% y en análisis más estrictos, para un nivel de confianza del 99%.

Podemos deducir, a la vista del cuadro, que puede rechazarse la hipótesis nula H_0 : "Pobla(I) no es causada por PIB(I)" en todos los casos y a un nivel de significancia del 1%, salvo para el caso de dos rezagos en que sí se rechazaría a un nivel de confianza del 95%, y que en ningún caso podemos rechazar la hipótesis nula H_0 : "PIB(I) no es causada por Pobla(I)" ya que los valores correspondientes de F no exceden los valores críticos de la distribución F para un nivel de confianza del 95%. La columna encabezada "p" (p value o probabilidad) muestra, precisamente, el grado de probabilidad de cumplimiento de la hipótesis nula, que no debiera rechazarse si tal valor p es superior a 0,05 (nivel de confianza del 95%) o a 0,01 (nivel de confianza del 99%).

A tenor de los datos y de su correspondiente interpretación podemos concluir que existe causalidad de PIB(I) hacia Pobla(I), pero no existe

causalidad de Pobl(I) hacia PIB(I), de forma que son los cambios en la variable PIB(I) los que actúan como causa y provocan efectos en la variable Pobl(I) que, sería, por ello mismo, la variable dependiente. Esta relación es especialmente manifiesta si consideramos un solo rezago o retardo, obteniendo valores de F que no dejan lugar a dudas: $F = 18,777$ permite rechazar la no causalidad de PIB(I) y $F=1,328$ no permite rechazar la no causalidad de la variable Pobl(I).

Mediante el contraste de Granger hemos podido detectar estadísticamente la dirección de la causalidad entre nuestras dos series básicas libres de tendencia, pero su utilidad, en relación con la validación de nuestra hipótesis nº 1, puede continuar aún si consideramos su validez para desechar la independencia de dos series. A este respecto la prueba de Granger (Gujarati, pág. 607) sugiere independencia cuando los conjuntos de coeficientes no son estadísticamente significativos en ambas regresiones, es decir, cuando no puede rechazarse ninguna de las dos hipótesis nulas analizadas para cada rezago.

En nuestro caso, y a un alto nivel de confianza, pudimos rechazar la hipótesis nula de no causalidad del PIB(I) destruyendo una de las condiciones para la independencia entre las series Pobl(I) y PIB(I), circunstancia que nos permite confirmar mediante el contraste de Granger nuestras estimaciones teóricas contenidas en la hipótesis nº 1: "La población no es una variable independiente". Es dependiente, al menos de la variable PIB, que actúa como causa y provoca cambios (efectos) en ella. Las limitaciones del análisis, la generalidad de las series, o la poca aplicabilidad de los métodos no restan valor al enunciado final, que no es otro que la ¡validación de la hipótesis nº 1, y ello porque hemos de considerar que aunque sólo estemos vislumbrando o percibiendo en la penumbra, cuanto vislumbramos apunta en la dirección enunciada: la población no es una variable independiente. Científicamente, mientras analíticamente no se perciba lo contrario, estamos legitimados para aceptar la dependencia (por no independencia) de la variable población.

3.2.3 Generalizando la no independencia Economía - Población

Aunque el propósito de la Tesis abarca el panorama español del siglo XX, es interesante, una vez establecida la “no independencia” en el ámbito español entre las variables “población” y “economía” (marcha de la economía simbolizada por el PIB), ampliar el estudio a diversos países para constatar si esta relación “no independiente” está ampliamente extendida o incluso generalizada, de forma que la realidad española no vendría a ser una excepción sino un elemento más de la regularidad observable entre la evolución de la población de un país y la marcha de su economía, entre las cuales existiría “asociación” o correlación significativa que nos llevaría a afirmar que ambas series de valores no son independientes.

Para extender, pues, el estudio de no independencia estadística, a otros países, hemos llevado a cabo la recopilación de datos sobre población y PIB de 5 países occidentales y del entorno de España: Francia, Bélgica, Alemania, Italia y Reino Unido. Para garantizar la homogeneidad con los resultados obtenidos en el caso español, la fuente de datos es la misma y aplicamos la misma metodología allí empleada: Contraste de independencia de Gujarati, para lo cual determinamos la tendencia por el método de Oskar Lange, obtenemos los residuos que serán nuestras series libres de tendencia, revisamos su normalidad mediante mediciones de asimetría, curtosis y contraste de Jarque-Bera, y calculamos el coeficiente de correlación de Pearson para establecer, en caso de correlación significativa, la no independencia de las variables.

La elección de los países responde, por una parte, a la proximidad geográfica, histórica y cultural de España. Todos ellos responderían a la categoría de países occidentales y capitalistas avanzados. Por otra, se debe también a la disponibilidad de datos, de mayor fiabilidad y homogeneidad, que abarcan todo el siglo XX, en concreto desde 1900 a 1994, y de la misma fuente que en el caso español, la obra “La economía mundial 1820-1992” de Angus Maddison y editado por la OCDE. No se trata de realizar un exhaustivo estudio mundial, pero este abanico de cinco países europeos puede reforzar la

conclusión obtenida para el caso español, o en su caso, señalar la singularidad española si sólo fuera significativa la no independencia entre economía y población para los valores de las variables españolas. Para cada país analizado mostraremos los datos, gráficas y resultados obtenidos, resaltando las conclusiones que cabe deducir de los mismos.

En el caso de las series del PIB pertenecientes a cada uno de los países, se trata de series temporales de datos anuales, valorados en dólares de 1990 y según el enfoque Geary-Khamis para permitir comparaciones multilaterales (está basado en los conceptos gemelos de paridad del poder adquisitivo y precios promedios internacionales de mercancías). La homogeneidad de las series, obtenidas por el mismo enfoque en todos los países, no ofrece dudas, siendo la Paridad del Poder Adquisitivo (PPA) de Geary-Khamis generalmente la más cercana a la de Paasche, la obtenida por el método de Fisher un poco más elevada y la de Laspeyres la más alta de todas. En términos generales la dispersión entre PPA's alternativas es más amplia mientras más bajo es el ingreso per cápita del país en cuestión, por lo que se obtienen resultados más fiables en comparaciones de países, si éstos poseen cierta similitud en su dimensión económica.

Las series de población en los cinco países analizados recogen los datos decenales de censos nacionales y las interpolaciones llevadas a cabo por la OCDE, de forma que se presentan en series anuales. Dado que en algunos países de los analizados se han efectuado durante el siglo XX temporales cambios de fronteras, el autor de los datos manejados, Angus Maddison, ha llevado a cabo los ajustes necesarios para que las cifras respondan a las fronteras de 1990 y la serie abarque poblaciones de los territorios actuales. Para mayor detalle véase la obra "La economía mundial 1820-1990" del ya citado Angus Maddison. Recordando las limitaciones de los datos anuales de población que recogíamos para el caso español, añade Angus Maddison que las cifras a frontera constante son un tanto anómalas, aunque es la mejor opción disponible si queremos ser congruentes con las series del PIB. En nuestro caso es además, totalmente necesario que ambas series de población y PIB estén referidas a fronteras idénticas.

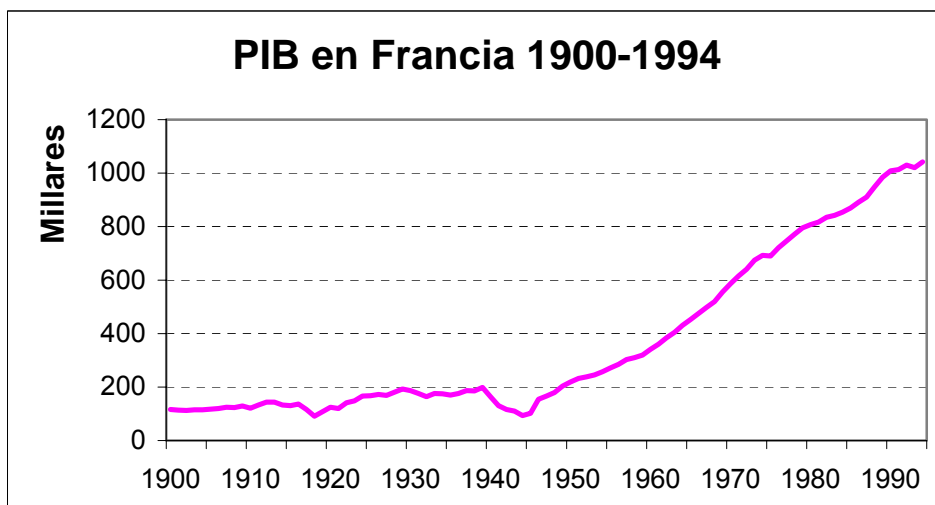
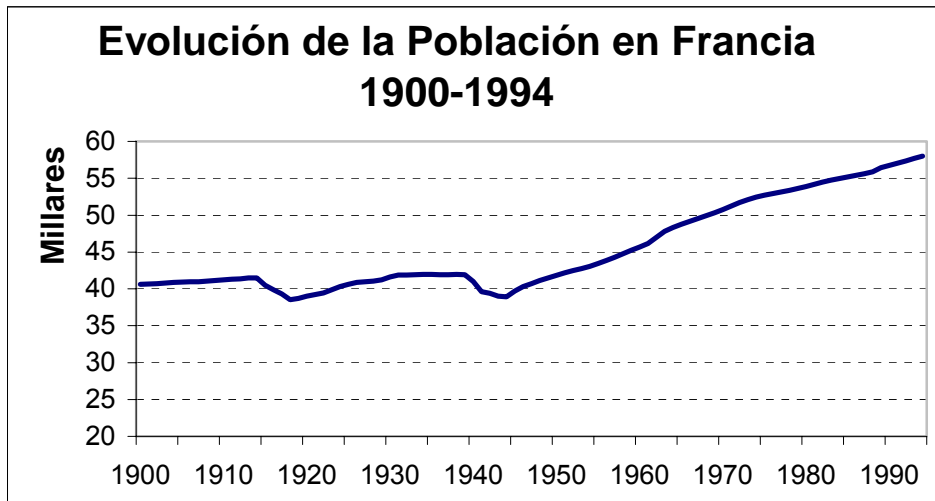
A) Estudio de la independencia Economía-Población en FRANCIA

Recogemos a continuación los datos correspondientes a Francia y para los años 1900-1994, población en miles de habitantes y PIB en miles de millones de dólares constantes de 1990.

Año	Población	PIB	Año	Población	PIB
1900	40598	115645	1948	41110	178906
1901	40640	113784	1949	41480	203237
1902	40713	111924	1950	41836	218409
1903	40786	114357	1951	42156	231862
1904	40859	115216	1952	42460	238017
1905	40890	117219	1953	42752	244887
1906	40942	119366	1954	43057	256766
1907	40942	124519	1955	43428	271508
1908	41046	123803	1956	43843	285248
1909	41109	128956	1957	44311	302423
1910	41224	121084	1958	44789	310008
1911	41307	132963	1959	45240	318882
1912	41359	143984	1960	45684	341353
1913	41463	143125	1961	46163	360102
1914	41476	132963	1962	46998	384147
1915	40481	130244	1963	47816	404757
1916	39884	136827	1964	48310	431092
1917	39288	115931	1965	48758	451702
1918	38542	91457	1966	49164	475318
1919	38700	107773	1967	49548	497502
1920	39000	124662	1968	49915	518685
1921	39240	119509	1969	50315	555038
1922	39420	140978	1970	50772	586812
1923	39880	148277	1971	51251	614721
1924	40310	166884	1972	51701	639768
1925	40610	167599	1973	52118	674404
1926	40870	172036	1974	52460	692724
1927	40940	168458	1975	52699	690434
1928	41050	180194	1976	52909	720920
1929	41230	192360	1977	53145	746396
1930	41610	186778	1978	53376	771586
1931	41860	175614	1979	53606	796061
1932	41860	164164	1980	53880	807081
1933	41890	175900	1981	54182	816814
1934	41950	174183	1982	54480	835563
1935	41940	169746	1983	54729	842147
1936	41910	176187	1984	54947	854599
1937	41930	186349	1985	55170	870199
1938	41960	185633	1986	55394	891095
1939	41900	198944	1987	55630	910417
1940	41000	164164	1988	55884	949204
1941	39600	129814	1989	56423	984985
1942	39400	116361	1990	56735	1008601
1943	39000	110492	1991	57050	1012895
1944	38900	93317	1992	57372	1030356
1945	39700	101189	1993	57696	1020051
1946	40290	153859	1994	58022	1042522
1947	40680	166740			

Fuente: Angus Maddison. La economía mundial 1820-1992

Las representaciones gráficas de los datos ya muestran una visible “asociación”, en particular la coincidencia de las brechas debidas a las contiendas mundiales, que siguen un mismo patrón en la serie de población y en la del PIB. A continuación se muestran ambos gráficos:

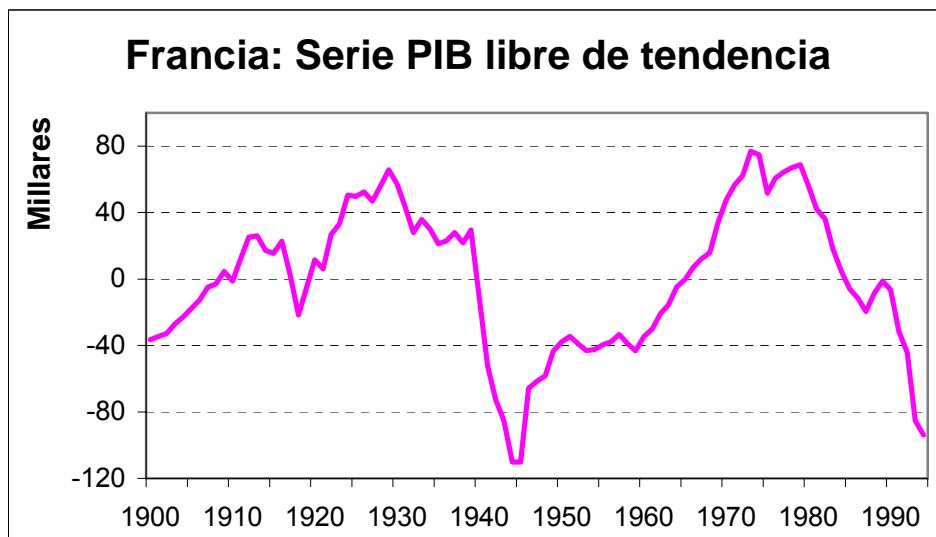
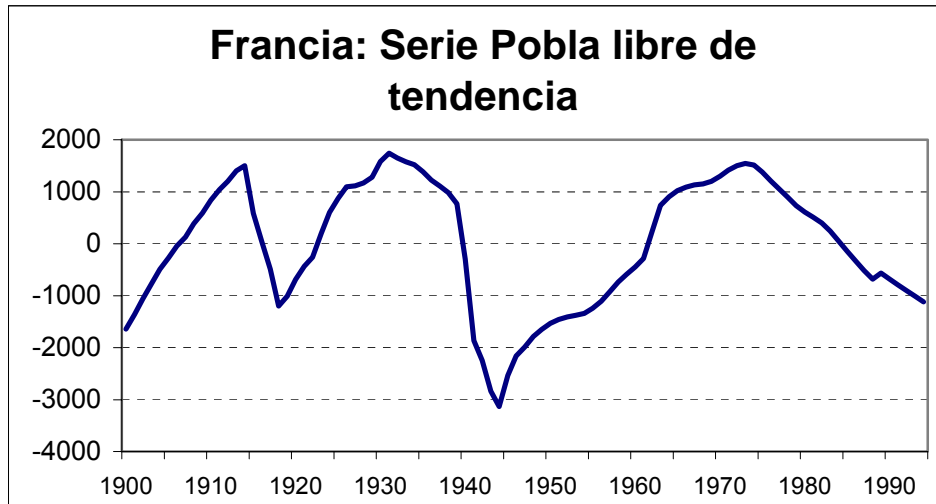


Es asimismo apreciable en ambos gráficos la tendencia creciente a partir de 1945, con mayor pendiente en el caso del PIB. Para depurar la serie y convertirla en estacionaria, tal como explicábamos en el caso español, estimamos en primer lugar la tendencia por el método de Oskar Lange, resultando ser la forma cúbica la que presenta mejor ajuste. Efectuando las regresiones

$$\text{Pobla} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

$$\text{PIB} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

Obtenemos como residuos las que serán las series libres de tendencia, cuyas representaciones gráficas permiten apreciar la existencia de patrones definidos (ciclos), lo cual no afecta a los resultados que pretendemos alcanzar.



Calculamos para ambas series los estadísticos “coeficiente de asimetría” y “Kurtosis”, así como el valor correspondiente al contraste de Jarque-Bera, obteniendo los valores siguientes

Serie: Población 1900-1994 libre de tendencia
 Simetría = -0,46 Kurtosis = 3,02 Contraste JB = 3,35

Serie: PIB 1900-1994 libre de tendencia
 Simetría = 0,334 Kurtosis = 2,63 Contraste JB = 2,30

En ambos casos los valores de asimetría y Kurtosis, por su proximidad a 0 y 3, respectivamente, permiten aceptar la afinidad con la distribución normal. En cuanto al contraste de Jarque Bera, más afinado, que contrasta la hipótesis nula H_0 : “La distribución es normal” mediante el estadístico JB que, asintóticamente, sigue una distribución chi-cuadrado con dos grados de libertad, no puede rechazarse, para un nivel de confianza del 95 %, para valores inferiores a 6 (las tablas muestran el valor 5,99). Por tanto, los valores obtenidos en las series libres de tendencia, correspondientes a la población y PIB de Francia, 3,35 y 2,30 respectivamente, no nos permiten rechazar la hipótesis nula de que las dos series pueden seguir la distribución normal.

El siguiente paso consiste en el cálculo de la covarianza o de la correlación entre las dos series para confirmar su valor distinto de cero. Mediante la conocida fórmula del coeficiente de correlación $r = S_{xy} / S_x \cdot S_y$ obtenemos: $r = 0,822$ valor que nos permite alegar que no existe independencia entre ambas series por su grado de alejamiento del valor nulo (recordemos que el valor de r puede oscilar entre -1 y +1). En consecuencia puede aceptarse la validez de nuestra hipótesis n° 1 que establecía la no independencia de la variable población también para el caso de Francia.

B) Estudio de la independencia Economía-Población en BÉLGICA

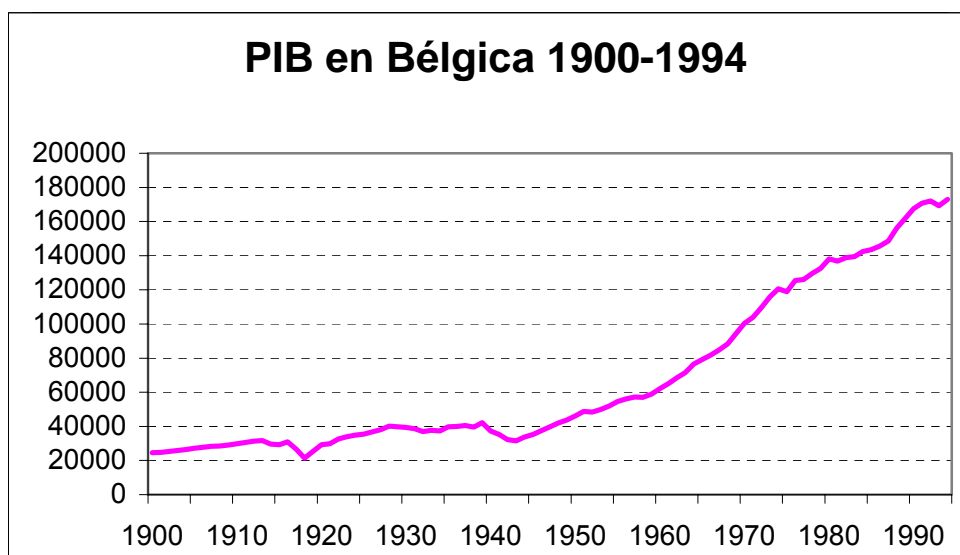
Mostramos a continuación los datos correspondientes a Bélgica y para los años 1900-1994, población en miles de habitantes y PIB en miles de millones de dólares constantes de 1990.

Año	Población	PIB	Año	Población	PIB
1900	6719	24537	1948	8557	42078
1901	6801	24759	1949	8614	43787
1902	6903	25266	1950	8640	46194
1903	6997	25835	1951	8679	48821
1904	7086	26500	1952	8731	48442
1905	7175	27260	1953	8778	49993
1906	7258	27830	1954	8820	52051
1907	7338	28242	1955	8869	54520
1908	7411	28527	1956	8924	56104
1909	7478	29065	1957	8989	57148
1910	7498	29825	1958	9053	57085
1911	7517	30521	1959	9104	58890
1912	7590	31250	1960	9154	62056
1913	7666	31661	1961	9184	65159
1914	7723	29657	1962	9218	68546
1915	7759	29300	1963	9283	71522
1916	7762	31001	1964	9367	76493
1917	7729	26622	1965	9448	79216
1918	7660	21453	1966	9508	81717
1919	7628	25306	1967	9557	84883
1920	7552	29287	1968	9590	88461
1921	7504	29793	1969	9613	94350
1922	7571	32706	1970	9638	100334
1923	7635	33877	1971	9673	104007

1924	7707	34986	1972	9711	109516
1925	7779	35524	1973	9742	115975
1926	7844	36727	1974	9772	120724
1927	7904	38088	1975	9801	118919
1928	7968	40083	1976	9818	125568
1929	8032	39735	1977	9845	126170
1930	8076	39355	1978	9872	129621
1931	8126	38658	1979	9837	132375
1932	8186	36917	1980	9847	138074
1933	8231	37708	1981	9853	136713
1934	8262	37392	1982	9856	138771
1935	8288	39703	1983	9855	139404
1936	8315	39988	1984	9855	142412
1937	8346	40526	1985	9858	143583
1938	8374	39608	1986	9862	145673
1939	8392	42299	1987	9870	148649
1940	8346	37265	1988	9921	156058
1941	8276	35302	1989	9938	161978
1942	8247	32263	1990	9967	167519
1943	8242	31515	1991	10005	170685
1944	8291	33731	1992	10025	172078
1945	8339	35365	1993	10045	169165
1946	8367	37455	1994	10045	173028
1947	8450	39703			

Fuente: Angus Maddison. La economía mundial 1820-1992

Y sus representaciones gráficas, ilustrativas de los avatares seguidos durante el siglo, son las siguientes:



Como puede observarse, ambas series se mueven en la misma dirección, con una perceptible “asociación” o paralelismo, de propensión creciente, con brechas en las fechas correspondientes a las dos guerras mundiales. Ya quedó apuntado que la propensión creciente es circunstancia

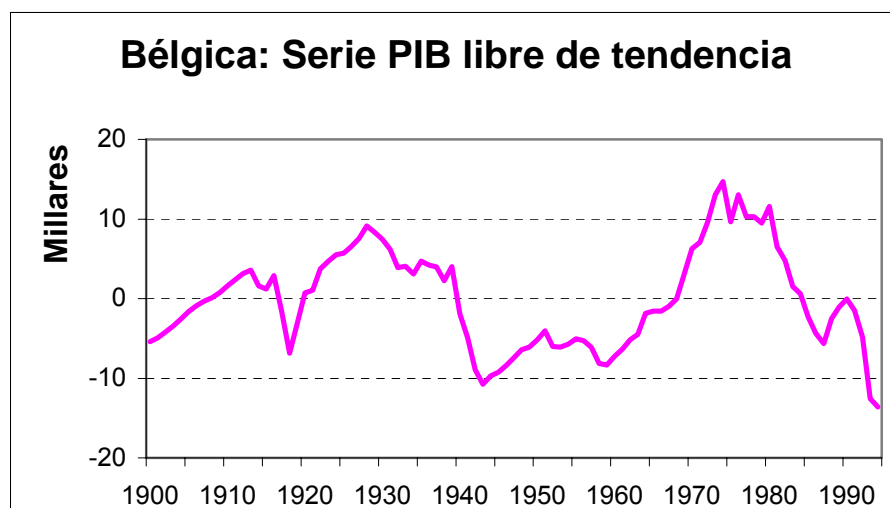
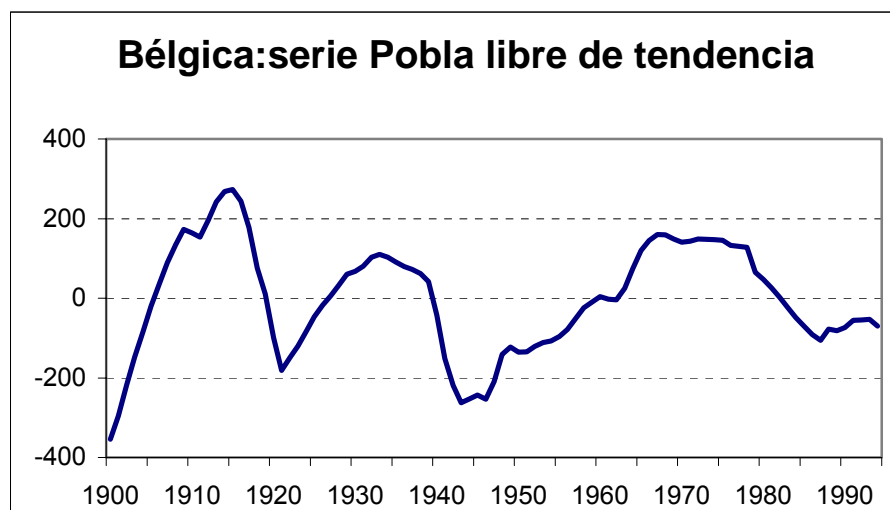
frecuente en información relacionada con series de tiempo económicas y que ha de alertarnos para evitar el problema de correlación espuria, puesto que en ese caso un alto valor de R cuadrado puede no reflejar la verdadera asociación entre ellas, sino simplemente la inclinación común en las dos series.

Precisamente, para hacer nuestras series estacionarias y evitar el peligro de correlación espuria recurrimos a la eliminación de las tendencias, en nuestro caso adoptando la forma cúbica que es la que produce, una vez llevado a cabo el método de Oskar Lange, el mejor ajuste. Para ello efectuamos dos regresiones según las siguientes ecuaciones:

$$\text{Pobla} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

$$\text{PIB} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

Los residuos obtenidos serán nuestras series libres de tendencia, en adelante Pobl(I) y PIB(I), las cuales son estacionarias y permiten que llevemos a cabo los cálculos de correlación para, una vez admitida su acercamiento a la distribución normal, establecer la independencia entre ellas si el coeficiente de correlación es cero. Veamos los gráficos de ambas series libres de tendencia:



Los estadísticos “coeficiente de asimetría” y “Kurtosis”, así como el valor correspondiente al contraste de Jarque-Bera, son los siguientes:

Serie: Población 1900-1994 libre de tendencia
 Simetría = -0,246 Kurtosis = 2,52 Contraste JB = 1,87

Serie: PIB 1900-1994 libre de tendencia
 Simetría = 0,205 Kurtosis = 2,43 Contraste JB = 1,99

Como vemos, no puede rechazarse la normalidad. El coeficiente de correlación entre ambas series nos da el valor $r = 0,597$ que, por su significatividad, nos permite afirmar que no hay independencia entre las series de población y economía (PIB) correspondientes a Bélgica.

C) Estudio de la independencia Economía-Población en ALEMANIA

El siguiente cuadro muestra los datos correspondientes a Alemania y para los años 1900-1994, población en miles de habitantes y PIB en miles de millones de dólares constantes de 1990.

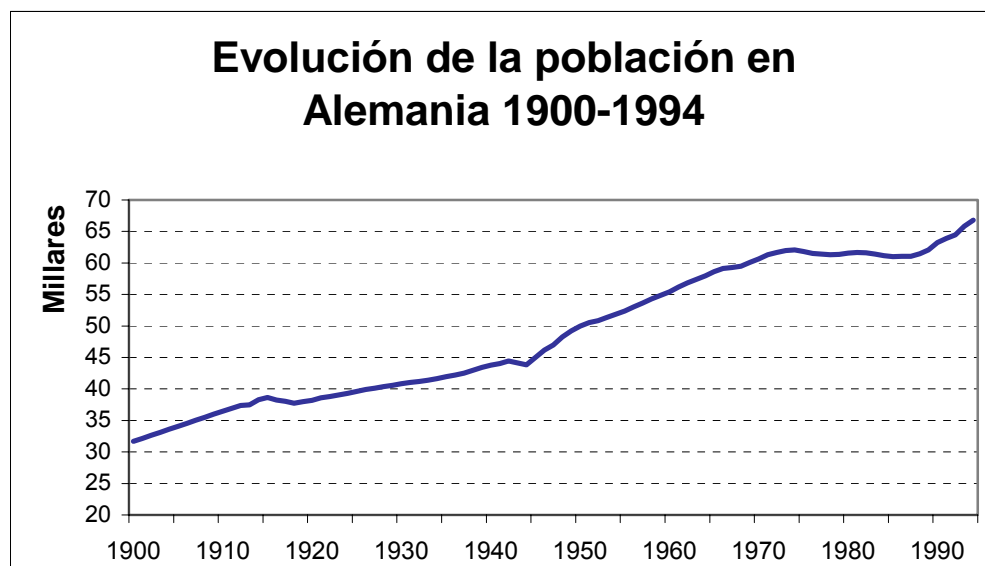
Año	Población	PIB
1900	31666	99227
1901	32134	96906
1902	32639	99227
1903	33126	104739
1904	33604	108946
1905	34078	111267
1906	34552	114604
1907	35038	119681
1908	35518	121712
1909	36001	124718
1910	36481	128676
1911	36928	133028
1912	37373	138830
1913	37483	145068
1914	38301	123598
1915	38637	117630
1916	38260	118521
1917	38063	118666
1918	37749	118956
1919	37966	104884
1920	38184	114024
1921	38610	126935
1922	38815	138105
1923	39070	114749
1924	39314	134333
1925	39608	149420
1926	39899	143327

Año	Población	PIB
1948	48251	153772
1949	49198	179159
1950	49983	213976
1951	50528	235011
1952	50859	256626
1953	51350	279256
1954	51880	300726
1955	52382	336848
1956	53008	362526
1957	53656	383851
1958	54292	400533
1959	54876	431723
1960	55433	469151
1961	56175	490331
1962	56837	512816
1963	57389	527033
1964	57971	562139
1965	58619	592314
1966	59148	609142
1967	59286	607256
1968	59500	640912
1969	60067	688639
1970	60651	723745
1971	61302	745506
1972	61672	777276
1973	61976	815138
1974	62054	817314

1927	40146	170020	1975	61829	805854
1928	40378	173937	1976	61531	848939
1929	40595	175968	1977	61400	873021
1930	40811	165233	1978	61327	899278
1931	41024	148405	1979	61359	936125
1932	41207	134623	1980	61566	946280
1933	41402	148695	1981	61682	947151
1934	41642	160155	1982	61638	938301
1935	41932	174662	1983	61423	954839
1936	42208	192941	1984	61175	981677
1937	42534	204546	1985	61024	1001551
1938	42990	220359	1986	61066	1025052
1939	43446	241103	1987	61077	1040284
1940	43792	242844	1988	61451	1079018
1941	44047	258221	1989	62063	1118041
1942	44417	261703	1990	63254	1181871
1943	44151	266926	1991	63889	1235546
1944	43809	273744	1992	64486	1254840
1945	45000	194682	1993	65817	1241059
1946	46190	115619	1994	66802	1275730
1947	46992	129836			

Fuente: Angus Maddison. La economía mundial 1820-1992

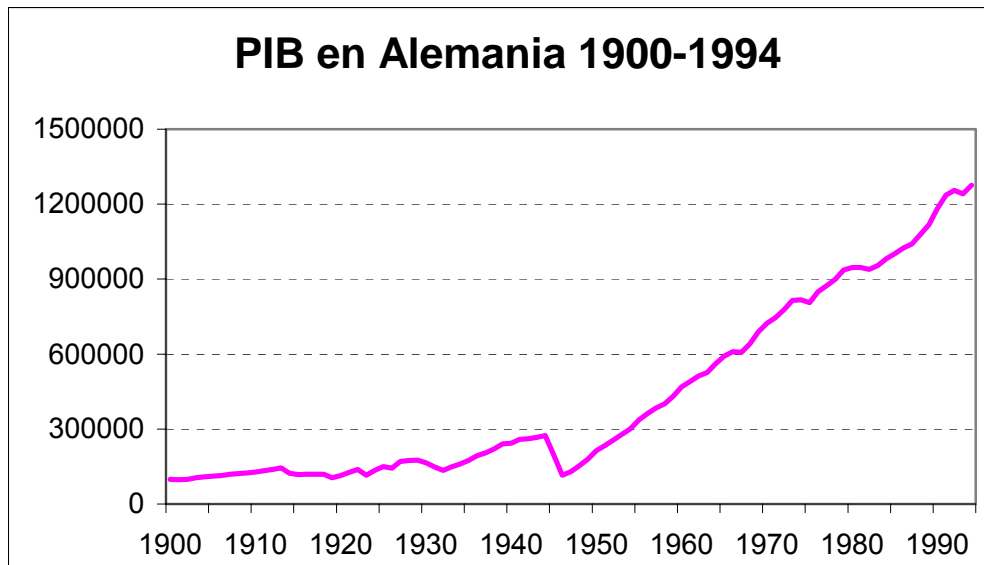
A continuación se muestran las representaciones gráficas de las dos series de datos, población y PIB en Alemania. Son destacables las sangrías que produjeron las dos guerras mundiales, con una severa caída en el caso del PIB tras la II Guerra, al tiempo que el “milagro alemán” queda patente en la vigorosa recuperación, de la economía y de la población, que tiene lugar una vez alcanzada la paz.



Siendo ambas series de propensión creciente hay, no obstante, tramos no concordantes, como puede apreciarse en los gráficos entre 1970 y 1990. En esa fase la población se estanca al nivel de 61 millones, mientras que el PIB prácticamente se duplica, pasando desde los 700.000 a los 1.200.000 millones de dólares constantes de 1900. Esta fase muestra poca “asociación”

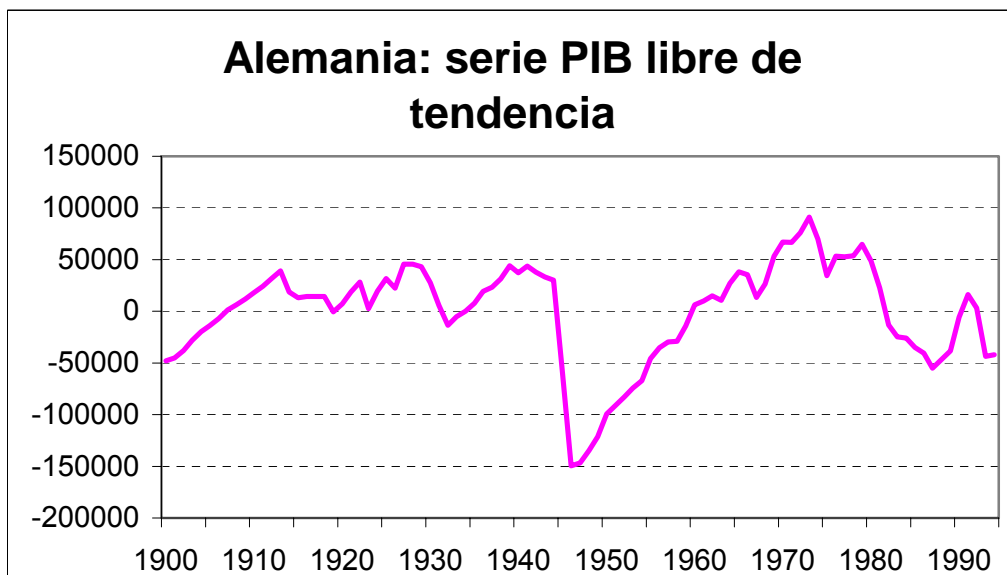
Nuevamente, para hacer nuestras series estacionarias y evitar el peligro de correlación espuria recurrimos a la eliminación de las tendencias, adoptando la forma cúbica una vez llevado a cabo el método de Oskar Lange. Para ello efectuamos dos regresiones según las siguientes ecuaciones:

$$\text{Pobla} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$



$$\text{PIB} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

Los residuos obtenidos serán nuestras series libres de tendencia, las cuales son estacionarias y permiten que llevemos a cabo los cálculos de correlación para, una vez admitida su acercamiento a la distribución normal,



establecer la independencia entre ellas si el coeficiente de correlación es cero.

Estas dos series libres de tendencia han perdido ya la propensión creciente, mostrando ciclos coincidentes a lo largo del recorrido, salvo en el último tramo desde 1980, en que parece desaparecer la semejanza entre los ciclos de población y economía. En cualquier caso el fragoroso desenvolvimiento histórico de Alemania durante el siglo XX podría calificarse de cualquier modo menos de apacible senda, moviéndose su población y su economía (PIB) con notable vigor en comparación con otros países europeos de su entorno, y sufriendo acentuadas caídas por las consecuencias bélicas, más profunda aquí que en esos mismo países europeos

Para el caso alemán, series de población y PIB libres de tendencia, los estadísticos “coeficiente de asimetría” y “Kurtosis”, así como el valor correspondiente al contraste de Jarque-Bera, son los siguientes:

Serie: Pobl(I) Población 1900-1994 libre de tendencia
Simetría = -0,236 Kurtosis = 2,35 Contraste JB = 3,04

Serie: PIB(I) PIB 1900-1994 libre de tendencia
Simetría = -1,03 Kurtosis = 1,84 Contraste JB = 16,44

La normalidad, que no puede rechazarse en la serie Pobl(I) al nivel de confianza del 95%, puede ser rechazada a ese nivel en la serie (PIB) ya que el valor del estadístico JB es superior a 6. Tanto el valor del coeficiente de simetría (-1,03, alejado del valor central 0) como el reducido valor de la Kurtosis (1,84, alejado del valor central 3) producen esta indefinición respecto a la distribución normal, si bien para nuestros fines aceptaremos la normalidad, siquiera sea con un menor grado de confianza respecto a los demostrados en anteriores series.

. El coeficiente de correlación entre ambas series libres de tendencia nos da el valor $r = 0,399$ que, al estar alejado del valor central cero, tiene la consideración de significativo, y es esta significatividad, la que nos permite afirmar que no hay independencia entre las series de población y economía (PIB) correspondientes a Alemania, de modo que también aquí quedaría respaldada nuestra hipótesis nº 1.

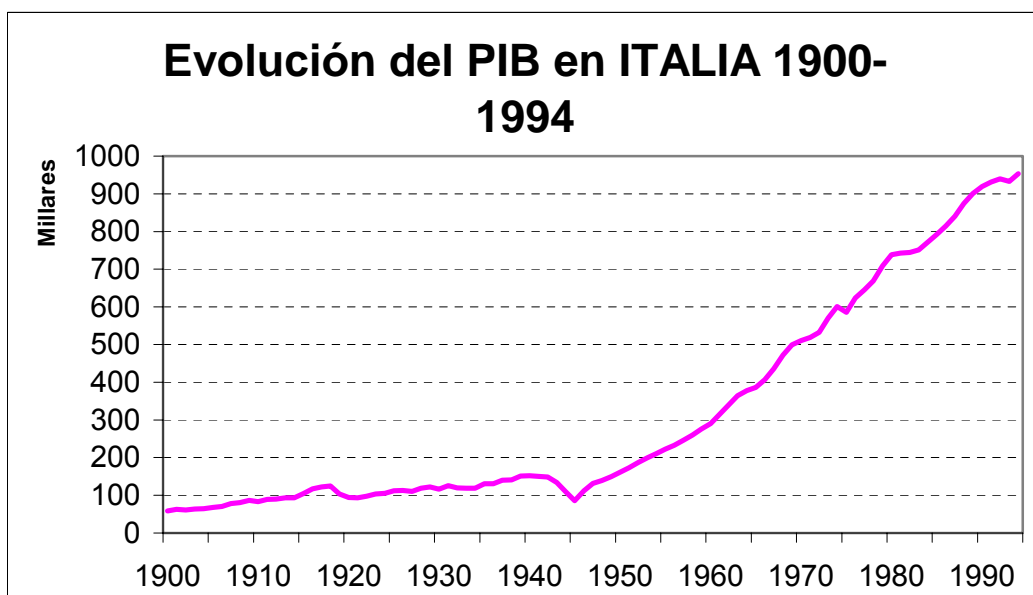
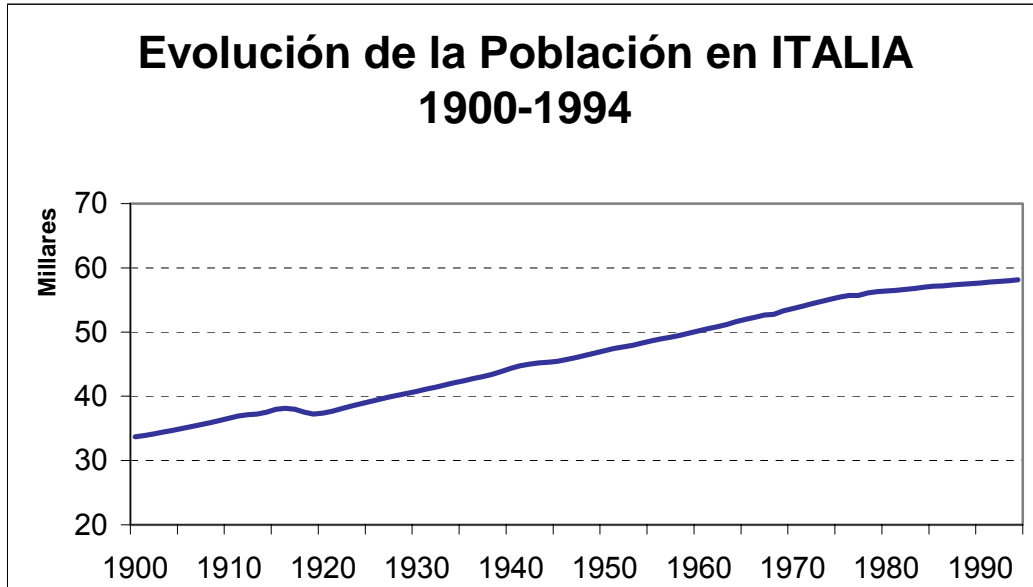
D) Estudio de la independencia Economía-Población en ITALIA

El siguiente cuadro muestra los datos correspondientes a Italia y para los años 1900-1994, población en miles de habitantes y PIB en miles de millones de dólares constantes de 1990.

Año	Población	PIB	Año	Población	PIB
1900	33672	58799	1948	46381	138967
1901	33877	62616	1949	46733	149227
1902	34166	60870	1950	47105	161351
1903	34436	63770	1951	47418	173476
1904	34715	64366	1952	47666	186421
1905	35011	67958	1953	47957	199899
1906	35297	70511	1954	48299	210270
1907	35594	78460	1955	48633	222475
1908	35899	80353	1956	48921	232562
1909	36213	86558	1957	49182	246325
1910	36572	83420	1958	49476	259449
1911	36917	88852	1959	49832	275630
1912	37150	89571	1960	50198	290574
1913	37248	93399	1961	50524	315053
1914	37526	93326	1962	50844	339591
1915	37982	104396	1963	51199	363813
1916	38142	117128	1964	51601	377977
1917	37981	122641	1965	51988	386481
1918	37520	124466	1966	52332	406657
1919	37250	103662	1967	52667	435584
1920	37398	94641	1968	52787	472048
1921	37691	93203	1969	53317	498996
1922	38086	98019	1970	53661	510200
1923	38460	103942	1971	54015	518457
1924	38810	104965	1972	54400	532466
1925	39165	111895	1973	54769	570200
1926	39502	113067	1974	55130	601208
1927	39848	110621	1975	55441	585330
1928	40186	118532	1976	55701	623810
1929	40469	122443	1977	55730	644825
1930	40791	116411	1978	56127	668514
1931	41132	125735	1979	56292	708617
1932	41431	119469	1980	56416	738598
1933	41753	118664	1981	56503	742707
1934	42093	119162	1982	56639	744295
1935	42429	130638	1983	56825	751487
1936	42750	130866	1984	56983	771661
1937	43068	139828	1985	57128	791741
1938	43419	140833	1986	57221	814904
1939	43865	151092	1987	57331	840402
1940	44341	152025	1988	57441	874586
1941	44734	150159	1989	57525	900271
1942	45004	148294	1990	57647	919511
1943	45177	134304	1991	57783	930999
1944	45290	109122	1992	57900	939685
1945	45442	85432	1993	58017	933147
1946	45725	111920	1994	58134	953602
1947	46040	131506			

Fuente: Angus Maddison. La economía mundial 1820-1992

Las representaciones gráficas de los datos anteriores, series de población y PIB en Italia durante el siglo XX, son las siguientes:



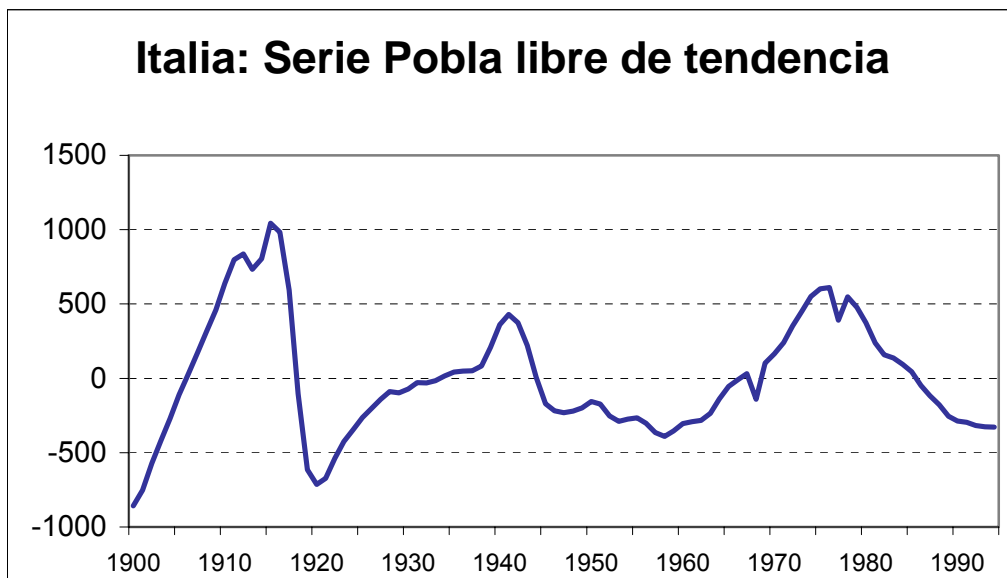
A la vista de los datos originales no se aprecia una ajustada asociación entre las líneas que representan las dos series correspondientes a Italia, más allá de una común tendencia creciente. Las caídas del PIB en los períodos bélicos no se corresponden con significativas caídas de la población y luego, a partir de 1950, la pendiente de la gráfica del PIB es muy superior a la de la línea de suave crecimiento de la población.

Tal y como hemos procedido en los casos anteriores, aplicando la metodología de eliminación de tendencia de Oskar Lange y adoptando la forma cúbica al ser la que produce el mejor ajuste, efectuamos dos regresiones según las siguientes ecuaciones:

$$\text{Pobla} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

$$\text{PIB} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

Los residuos obtenidos serán nuestras series libres de tendencia, las cuales son estacionarias y permiten que llevemos a cabo los cálculos de correlación para el estudio de la presunta independencia. Estas son las representaciones gráficas de las series una vez eliminada la tendencia.

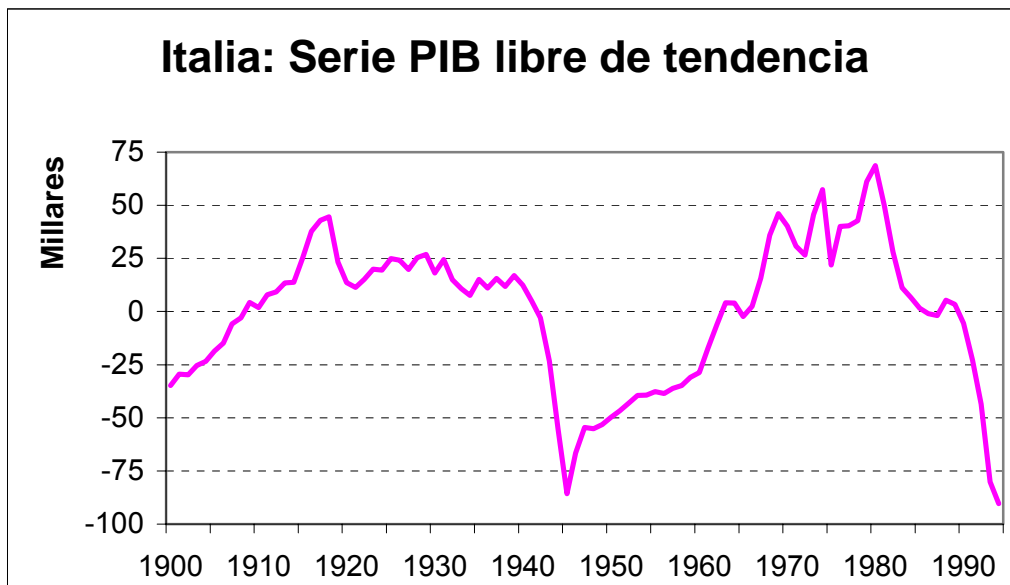


Para estas series (Italia) de población y PIB libres de tendencia, los estadísticos “coeficiente de asimetría”, “Kurtosis”, y contraste Jarque-Bera, presentan estos valores:

Serie: Pobl(l) Población 1900-1994 libre de tendencia
Simetría = 0,52 Kurtosis = 3,03 Contraste JB = 4,27

Serie: PIB(l) PIB 1900-1994 libre de tendencia
Simetría = -0,53 Kurtosis = 2,92 Contraste JB = 4,59

En ambos casos la normalidad, no puede rechazarse al nivel de confianza del 95%, ya que el valor del estadístico JB es inferior a 6.



. El coeficiente de correlación entre ambas series libres de tendencia nos da el valor $r = 0,502$, cuya significatividad, nos permite afirmar que no hay independencia entre las series de población y economía (PIB) correspondientes a Italia, respaldando nuestra hipótesis nº 1.

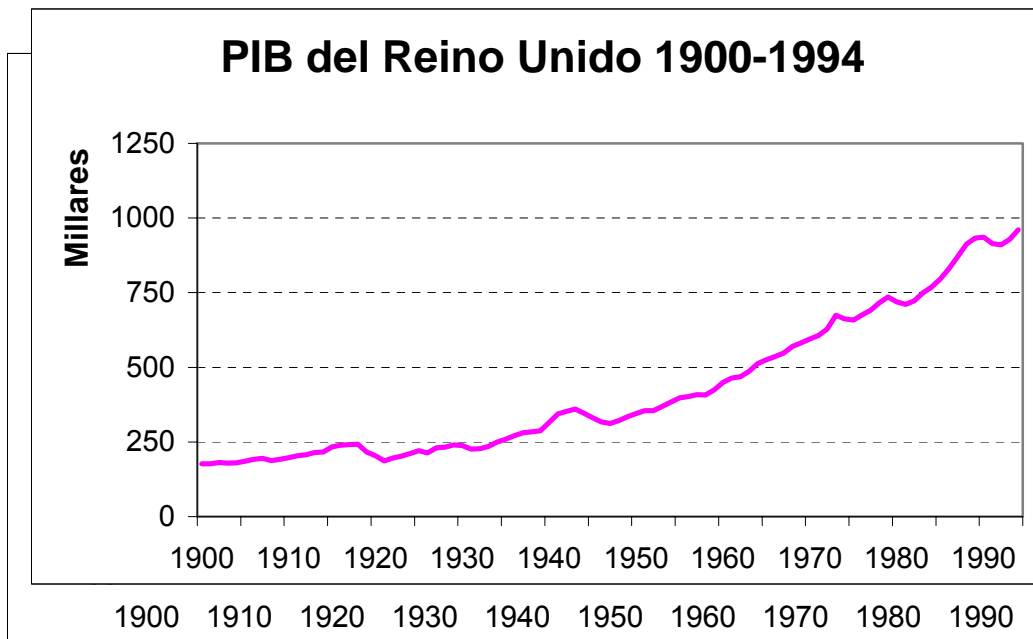
E) Estudio de la independencia Economía-Población en el REINO UNIDO

El siguiente cuadro muestra los datos correspondientes al Reino Unido para los años 1900-1994, población en miles de habitantes y PIB en miles de millones de dólares constantes de 1990.

Año	Población	PIB	Año	Población	PIB
1900	38426	176504	1948	50014	322125
1901	38784	176504	1949	50312	334135
1902	39115	181008	1950	50363	344859
1903	39445	179078	1951	50574	355153
1904	39786	180150	1952	50737	354509
1905	40131	185512	1953	50880	368450
1906	40486	191731	1954	51066	383462
1907	40837	195377	1955	51221	397402
1908	41199	187442	1956	51430	402335
1909	41568	191731	1957	51657	408769
1910	41938	197736	1958	51870	407911
1911	42267	203527	1959	52157	424425
1912	42424	206529	1960	52373	448874
1913	42622	214464	1961	52807	463672
1914	42996	216609	1962	53292	468390
1915	43268	233981	1963	53625	486405
1916	43430	239128	1964	53991	512141
1917	43523	241272	1965	54350	525437
1918	43487	242774	1966	54643	535517
1919	43449	216394	1967	54959	547527
1920	43718	203312	1968	55214	569832
1921	44072	186798	1969	55461	581627
1922	44372	196449	1970	55632	594924
1923	44596	202669	1971	55907	606719
1924	44915	211033	1972	56079	627737
1925	45059	221327	1973	56210	674061
1926	45232	213177	1974	56224	662695
1927	45389	230335	1975	56215	657762
1928	45578	233123	1976	56206	675777
1929	45672	239985	1977	56179	691433
1930	45866	238270	1978	56167	715667
1931	46074	226045	1979	56227	735827
1932	46335	227761	1980	56314	719528
1933	46520	234409	1981	56379	710306
1934	46666	249851	1982	56335	722530
1935	46868	259502	1983	56377	749767
1936	47081	271297	1984	56488	768425
1937	47289	280734	1985	56618	795233
1938	47494	284165	1986	56763	829548
1939	47991	286953	1987	56930	869009
1940	48226	315691	1988	57065	912331
1941	48216	344430	1989	57236	932276
1942	48400	353008	1990	57411	935922
1943	48789	360729	1991	57649	914690
1944	49016	346574	1992	57848	910401
1945	49182	331347	1993	58349	928630
1946	49217	316978	1994	58702	961014
1947	49519	312260			

Fuente: Angus Maddison. La economía mundial 1820-1992

Los siguientes gráficos corresponden a los datos anteriores:

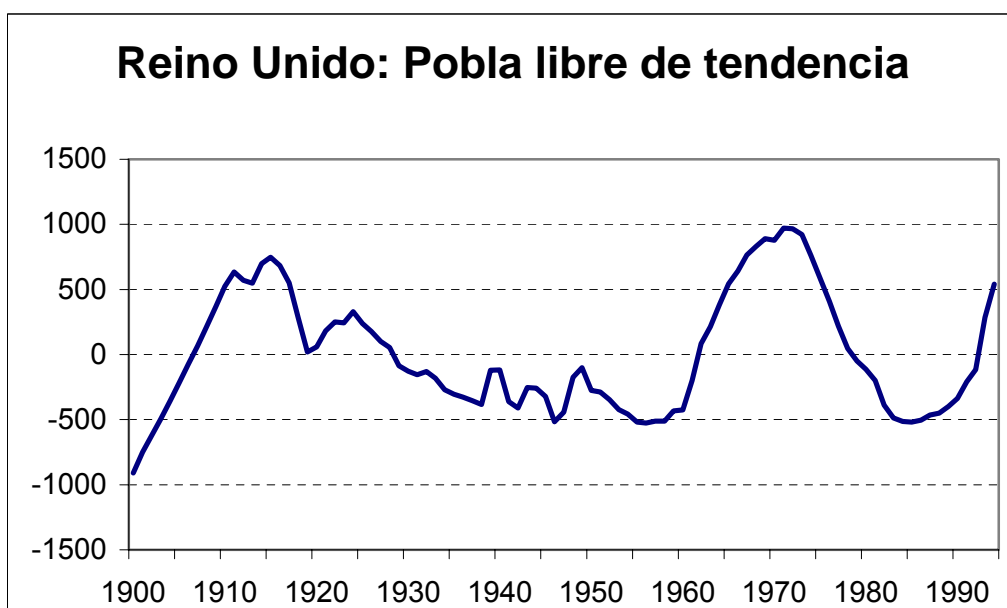


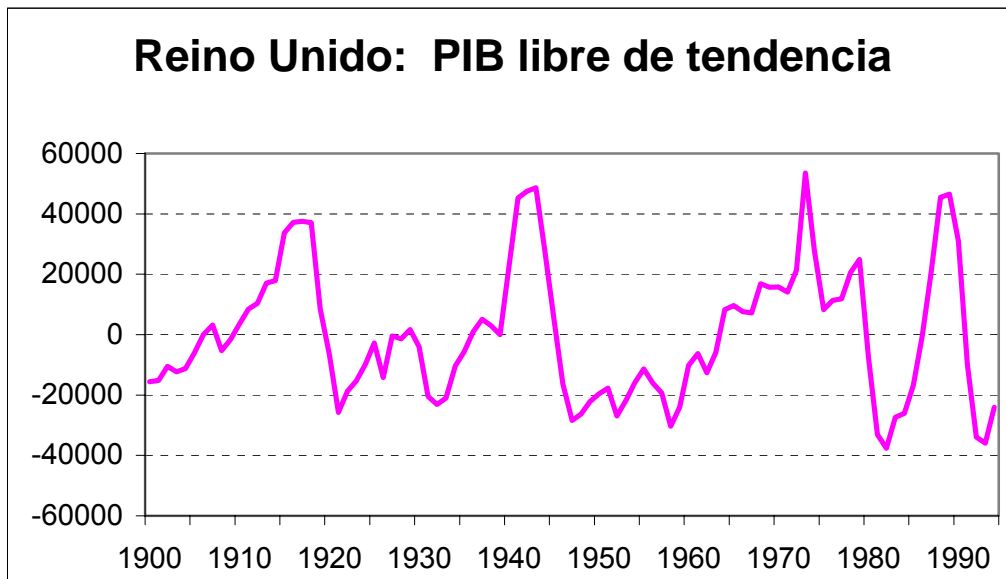
Nuevamente aplicamos la metodología de eliminación de tendencia de Oskar Lange, adoptando la forma cúbica al ser la que produce el mejor ajuste, para lo que efectuamos dos regresiones según las siguientes ecuaciones:

$$\text{Pobla} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

$$\text{PIB} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + V$$

Los residuos obtenidos son nuestras series libres de tendencia, las cuales son estacionarias y permiten que llevemos a cabo los cálculos de correlación para el estudio de la presunta independencia. He aquí sus gráficas:





Es posible deducir, de la propia visión de las gráficas, la estacionariedad que han alcanzado nuestras series de población y PIB, correspondientes al Reino Unido, libres de tendencia. Una vez efectuados los cálculos, los estadísticos “coeficiente de asimetría”, “Kurtosis”, y contraste Jarque-Bera, presentan estos valores:

Serie: Pobl(l) Población 1900-1994 libre de tendencia
 Simetría = 0,50 Kurtosis = 2,53 Contraste JB = 4,47

Serie: PIB(l) PIB 1900-1994 libre de tendencia
 Simetría = 0,57 Kurtosis = 2,72 Contraste JB = 4,59

En ambos casos la hipótesis nula $H(0)$ “la serie sigue la distribución normal” no puede rechazarse al nivel de confianza del 95%, ya que el valor del estadístico JB es inferior a 6.

. Calculando el coeficiente de correlación entre ambas series libres de tendencia obtenemos el valor $r = 0,389$, significativo por su alejamiento del cero y que nos permite afirmar que no hay independencia entre las series de población y economía (PIB) correspondientes al Reino Unido; en su lugar cierto grado de asociación está presente y demuestra que la evolución de ambas magnitudes ha estado relacionada, respaldando nuestra hipótesis nº 1.

Una vez realizado el estudio sobre la posible independencia estadística entre las series de población y economía (PIB) en los cinco países europeos seleccionados, procede deducir conclusiones de conjunto, a lo cual ayuda el siguiente cuadro de los coeficientes de correlación ya calculados:

País	Coefficiente de correlación entre series de Población y PIB libres de tendencia
FRANCIA	0,822
BÉLGICA	0,597
ALEMANIA	0,399
ITALIA	0,502
REINO UNIDO	0,389
ESPAÑA	0,715

Todos los coeficientes de correlación son significativos, indicadores de relación positiva entre las variables, de una asociación que contradice la presunta independencia entre ellas. Con mayor o menor grado, desde el 0,822 de Francia, muy significativo, al 0,389 del Reino Unido, ocupando el segundo lugar el valor obtenido para España: 0,715

Estos valores refuerzan cuanto decíamos para el caso español: por el alejamiento de estos coeficientes de correlación del valor nulo puede aceptarse, en todos los casos, la validez de la hipótesis nº 1 que establecía la no independencia de la variable población. Recordemos asimismo que no hemos podido rechazar la normalidad de las series de población y PIB libres de tendencia, al nivel de confianza del 95%, salvo en el caso del PIB alemán en que estábamos en una zona de menor confianza. La normalidad de las series y la significatividad de los coeficientes de correlación permiten afirmar que el caso español no es una excepción, antes bien corrobora un extendido o generalizado panorama de no independencia de la variable población.