

Replicación y Extensión de Pavia et al. (2017): Evaluación de la Calidad de las Añadas de la Contabilidad Nacional Trimestral Española con Análisis del COVID-19

Manuel A. Hidalgo-Pérez*
Universidad Pablo de Olavide

Leandro Navarro Pablo**
AIReF

15 de julio de 2025

Resumen

La fiabilidad de las estimaciones económicas preliminares es crucial para la toma de decisiones y la investigación. Este estudio realiza una doble contribución al análisis de la calidad de los datos de la Contabilidad Nacional Trimestral (CNTR) en España, tomando como referencia el trabajo seminal de Pavia et al. (2017). Primero, se lleva a cabo una replicación metodológica exhaustiva de su análisis para el período original 2005-2016, validando sus hallazgos clave, como la mayor precisión de las estimaciones del cuarto trimestre para las tasas de crecimiento interanual. Segundo, el análisis se extiende hasta 2025, lo que permite evaluar el comportamiento de los errores de revisión durante episodios de estrés económico sin precedentes. Nuestros resultados revelan una heterogeneidad notable en el impacto de las crisis; mientras que la pandemia de COVID-19 amplificó los errores de revisión en un factor de 2,94, este impacto fue superado por el de la crisis de deuda soberana (3,59). Al implementar todo el análisis en código Python de acceso abierto, no solo garantizamos la reproducibilidad total del estudio, sino que también ofrecemos una herramienta robusta para futuras investigaciones.

Palabras clave: Contabilidad Nacional, Revisiones de Datos, Replicación, COVID-19, Estadísticas Económicas
Códigos JEL: C82, E01, E32

*Universidad Pablo de Olavide, Sevilla. Email: mhidper@upo.es

**Autoridad Independiente de Responsabilidad Fiscal (AIReF). Email: lnavarro@airef.es

1. Introducción

La calidad y fiabilidad de las estadísticas económicas oficiales constituyen pilares fundamentales de la política económica basada en evidencia y la investigación académica. Entre estas estadísticas, los datos de Contabilidad Nacional Trimestral (CNTR) desempeñan un papel particularmente crucial en el seguimiento económico a corto plazo y la formulación de políticas. Sin embargo, el trade-off inherente entre puntualidad y precisión en la compilación de la CNTR significa que las estimaciones iniciales están sujetas a revisiones posteriores cuando se dispone de información más completa.

El trabajo seminal de Pavia et al. [2017] proporcionó un análisis comprensivo de los patrones de revisión en los datos de la CNTR española, estableciendo importantes referencias para entender los sesgos sistemáticos y los patrones estacionales en las estimaciones tempranas. Su estudio, que abarcó el período 2005-2016, reveló características significativas sobre la estructura temporal de los errores de revisión, particularmente la superior precisión de las estimaciones del cuarto trimestre para las tasas de crecimiento interanuales.

Partiendo de lo anterior, este trabajo persigue dos objetivos principales. Primero, proporcionamos una replicación metodológica completa de Pavia et al. (2017), implementando su marco analítico. Segundo, extendemos su análisis para incorporar datos hasta 2025, permitiéndonos examinar patrones de revisión durante disrupciones económicas sin precedentes, particularmente la pandemia del COVID-19.

Nuestra replicación confirma la robustez de los hallazgos originales mientras que nuestra extensión revela nuevas cuestiones sobre el comportamiento de los errores de revisión durante estrés económico extremo. El período COVID-19 muestra factores de amplificación de errores de revisión de casi 3x comparado con períodos normales, aunque notablemente menores que los observados durante la crisis europea de deuda soberana.

2. Revisión de Literatura y Marco Teórico

2.1. Literatura sobre Revisiones de Cuentas Nacionales

El análisis de revisiones de cuentas nacionales tiene una rica tradición académica que se remonta a Young [1993]. Una de las cuestiones fundamentales es que las estimaciones tempranas de agregados económicos enfrentan un trade-off inherente entre puntualidad y precisión, llevando a patrones de revisión sistemáticos que pueden ser analizados y potencialmente predichos.

La metodología de análisis de datos en tiempo real, popularizada en el contexto estadounidense por investigadores como Croushore y Stark, trata cada publicación de datos como una fotografía del conjunto de información disponible en un momento específico del tiempo. Este enfoque es crucial porque permite a los analistas recrear el entorno informativo al que se enfrentaban los responsables políticos en el pasado, evitando así juicios anacrónicos basados en datos finales que no estaban disponibles cuando se tomaron las decisiones.

El marco teórico para entender las revisiones fue formalizado por Pavia et al. [2017], quienes adaptaron el enfoque de vintages de Young al contexto español. Su innovación clave fue el análisis sistemático de patrones de revisión a través de trimestres, revelando que:

$$\text{Precisión} = f(\text{Estimación Inicial} - \text{Estimación Final}) \quad (1)$$

donde la precisión de las estimaciones iniciales varía sistemáticamente con el timing estacional de las publicaciones de datos y el entorno de información en el momento de la estimación.

2.2. Evidencia Internacional de Patrones de Revisión

La investigación internacional ha revelado patrones de revisión notablemente consistentes entre economías desarrolladas. El análisis del Banco Central Europeo de 2023 sobre Alemania, Francia, Italia y España encontró que las agencias estadísticas minimizan sistemáticamente las revisiones del PIB agregado mientras permiten un reequilibrio significativo entre los componentes del gasto. Este comportamiento refleja las preferencias institucionales para mantener credibilidad a través de cifras principales estables.

Alemania muestra un sesgo positivo único (0.051 puntos porcentuales) en las revisiones totales del PIB, mientras que Francia e Italia no exhiben sesgos agregados significativos. Sin embargo, todos los países muestran correlaciones negativas fuertes (50 %) entre las revisiones del comercio neto y las existencias, sugiriendo que las existencias sirven como un "elemento de equilibrio" para compensar las revisiones en otros componentes.

La Oficina Australiana de Estadísticas reporta revisiones promedio del PIB trimestral de 0.19 puntos porcentuales después de tres años, con las medidas de ingreso y gasto mostrando revisiones mayores que las medidas de producción. El desempeño de Australia supera la mediana de la OCDE, mientras que la Oficina Nacional de Estadísticas del Reino Unido documenta magnitudes de revisión similares (0.14 puntos porcentuales) pero patrones temporales diferentes.

2.3. Innovaciones Metodológicas Post-2017

El campo ha evolucionado dramáticamente más allá de las medidas básicas de Error Absoluto Medio (EAM) y sesgo, incorporando técnicas econométricas sofisticadas y de aprendizaje automático. El análisis de datos de añadas utilizando el marco de Mankiw-Shapiro distingue entre "noticias" (nueva información) y ruido (errores de medición), revelando que la mayoría de las revisiones reflejan actualizaciones genuinas de información en lugar de inadecuaciones estadísticas.

Las aplicaciones de aprendizaje automático han mostrado particular promesa, con métodos de conjunto que combinan múltiples modelos demostrando rendimiento superior durante períodos estables. Random Forest, XG-Boost y las redes neuronales capturan patrones de revisión no lineales que los métodos tradicionales no detectan, aunque tienen dificultades con eventos sin precedentes como COVID-19.

Los enfoques de series temporales utilizando modelos ARIMA y Autorregresión Vectorial (VAR) capturan dependencias temporales en los patrones de revisión, mientras que los métodos econométricos estructurales incluyendo modelos de espacio de estados con filtrado de Kalman proporcionan seguimiento de revisiones en tiempo real.

2.4. Patrones Estacionales en Errores de Revisión

Pavia et al. [2017] hipotizaron y confirmaron que las estimaciones del cuarto trimestre exhibirían precisión superior debido a:

- Horizonte de predicción más corto (más cercano a datos de cuentas anuales)
- Información más completa de trimestres precedentes
- Mayor disponibilidad de datos de fuentes administrativas

Esta hipótesis se formaliza en la expectativa de que $EAM_{Q4} < EAM_{Q3}$, donde EAM denota el Error Absoluto Medio.

2.5. Crisis Económicas y Degradación de la Calidad Estadística

La investigación sobre patrones de revisión a través de ciclos económicos revela que las revisiones son mayores y más persistentes durante las recesiones, con períodos de crisis mostrando patrones sistemáticos donde las estimaciones iniciales subestiman la severidad de la desaceleración. La crisis financiera de 2008 demostró esto dramáticamente, con las estimaciones del PIB del Q4 2008 revisadas de -3.8 % a -8.9 % a medida que se hizo aparente la magnitud completa de la crisis.

Las restricciones de datos en tiempo real resultan particularmente agudas durante períodos de crisis, afectando tanto la efectividad de la política monetaria como fiscal. La investigación extensa sobre la estimación de la regla de Taylor muestra que las recomendaciones de política difieren sustancialmente cuando se basan en datos en tiempo real versus datos revisados [Orphanides, 2001].

2.6. Literatura sobre Crisis Económicas y Calidad Estadística

Estudios recientes han comenzado a examinar cómo las crisis económicas afectan la calidad de las estadísticas oficiales. Aruoba [2008] documentó que los datos económicos no se comportan de manera uniforme durante períodos de estrés, mientras que Croushore [2011] enfatizó la importancia del análisis en tiempo real durante episodios de incertidumbre elevada.

La literatura sobre incertidumbre económica, particularmente Bloom [2009] y Baker et al. [2016], ha demostrado que los shocks de incertidumbre tienen efectos persistentes sobre la economía real. Nuestro trabajo contribuye a esta literatura examinando cómo la incertidumbre se manifiesta en la calidad de las estimaciones estadísticas oficiales.

La pandemia de COVID-19 aceleró la investigación en metodologías de revisión específicas para crisis, con el desacuerdo de los pronosticadores sobre el crecimiento del PIB aumentando ocho veces para Estados Unidos y veinte veces para el Reino Unido. Esta incertidumbre sin precedentes impulsó el desarrollo de indicadores de alta frecuencia mejorados y procedimientos de ajuste estacional modificados para manejar la volatilidad extrema.

2.6.1. La Macroeconomía de la Incertidumbre y la Microestructura de las Revisiones de Datos

La fiabilidad de las estadísticas económicas oficiales es la piedra angular sobre la que se construyen tanto la política económica basada en la evidencia como la investigación académica rigurosa. Sin embargo, la producción de estos datos se enfrenta a un compromiso fundamental entre la puntualidad y la precisión, lo que da lugar a un proceso de revisión continuo en el que las estimaciones iniciales se actualizan a medida que se dispone de información más completa. Este proceso de revisión, lejos de ser un mero ruido estadístico, revela profundas verdades sobre la naturaleza de la medición económica, especialmente durante períodos de elevada tensión e incertidumbre. El presente análisis desarrolla un marco teórico integral que conecta cuatro pilares de la literatura económica contemporánea para contextualizar y dar fundamento a la investigación sobre la calidad de las estadísticas de la Contabilidad Nacional.

Este marco se estructura en torno a las contribuciones seminales de Aruoba [2008], Croushore [2011], Bloom [2009] y Baker et al. [2016]. En conjunto, estos trabajos trazan una narrativa causal que va desde la naturaleza problemática de las revisiones de datos, pasando por la respuesta metodológica adecuada, hasta la teoría económica que explica por qué la incertidumbre degrada tanto la actividad económica real como la calidad de su medición estadística. El Cuadro 1 resume las contribuciones fundamentales que estructuran este informe.

Cuadro 1: Contribuciones Fundamentales a la Literatura sobre Revisiones de Datos e Incertidumbre

Autor(es)	Pregunta de Investigación Primaria	Contribución Clave	Relevancia Directa para este Estudio
Aruoba [2008]	¿Cuáles son las propiedades estadísticas de las revisiones de datos macroeconómicos?	Establece que las revisiones de datos “no se comportan bien” (son sesgadas, grandes y predecibles), lo que significa que las publicaciones iniciales no son pronósticos racionales.	Proporciona la teoría fundamental de por qué existen errores de revisión y por qué son sistemáticos, sentando las bases del problema a investigar.
Croushore [2011]	¿Cómo deben los economistas manejar las revisiones de datos en sus análisis?	Establece el paradigma de la “añada de datos en tiempo real” (real-time data vintage) como la metodología correcta para un análisis robusto, especialmente para la evaluación de políticas.	Proporciona la justificación metodológica para el análisis de diferentes añadas de datos (por ejemplo, A0 vs. DEF) que se realiza en el estudio de referencia.
Bloom [2009]	¿Cuál es el impacto de un shock de incertidumbre puro en la macroeconomía?	Desarrolla un modelo estructural que muestra que los shocks de incertidumbre, a través del canal de las opciones reales, provocan que las empresas pausen la inversión y la contratación, lo que lleva a recesiones agudas y recuperaciones rápidas.	Proporciona la teoría económica central que explica <i>por qué</i> los errores de revisión deberían amplificarse durante las crisis (períodos de alta incertidumbre).
Baker et al. [2016]	¿Cómo podemos medir la incertidumbre de la política económica y cuáles son sus efectos?	Desarrolla el índice de Incertidumbre de Política Económica (EPU) basado en texto y demuestra que predice caídas en la inversión, la producción y el empleo.	Ofrece un precedente empírico y un marco de medición para vincular la incertidumbre con los resultados económicos reales, que este estudio vincula con los resultados estadísticos.

2.6.2. El Desafío Fundamental: La Naturaleza “Mal Comportada” de los Datos Económicos

La investigación sobre la calidad de los datos económicos parte de un hecho ineludible: la mayoría de las estadísticas macroeconómicas, especialmente las de alta frecuencia como el Producto Interior Bruto (PIB) trimestral, están sujetas a revisiones. Este proceso surge del compromiso inherente entre la necesidad de disponer de información oportuna para la toma de decisiones y la realidad de que la recopilación de datos exhaustivos lleva tiempo. Las agencias estadísticas publican estimaciones preliminares basadas en información incompleta para satisfacer la demanda de los responsables políticos y los mercados, y posteriormente las actualizan a medida que llegan datos más completos y fiables, como los censos o los registros administrativos. La pregunta crucial, que permaneció subestimada durante décadas, es si estas revisiones son benignas, es decir, si se comportan como un ruido blanco aleatorio y no sistemático.

El trabajo seminal de Aruoba [2008], “Data Revisions Are Not Well Behaved”, proporcionó una respuesta contundente y negativa a esta pregunta, alterando fundamentalmente la percepción de la comunidad académica sobre la fiabilidad de los datos económicos iniciales. Analizando un amplio conjunto de variables macroeconómicas clave de Estados Unidos, Aruoba documentó de manera sistemática que las revisiones no satisfacen las “propiedades

estadísticas deseables simples” que se esperarían si fueran meros errores de medición aleatorios. Esta conclusión, a menudo resumida con la elocuente frase de que las revisiones “no se comportan bien”, se basa en tres hallazgos empíricos interconectados.

En primer lugar, las revisiones no tienen una media de cero. Este hallazgo de un sesgo sistemático es de suma importancia, ya que indica que las estimaciones iniciales publicadas por las agencias estadísticas no son, en promedio, una estimación insesgada de las cifras finales. Si las revisiones fueran puramente aleatorias, los errores positivos y negativos se cancelarían a lo largo del tiempo, resultando en una media nula. El hecho de que no lo hagan sugiere la presencia de un sesgo estructural en el proceso de estimación inicial.

En segundo lugar, la magnitud de estas revisiones es económicamente significativa. Aruoba [2008] demuestra que las revisiones son “bastante grandes en comparación con las variables originales”. Esto significa que las diferencias entre la primera estimación y la cifra final no son meras curiosidades estadísticas, sino que son lo suficientemente grandes como para alterar la evaluación del estado de la economía y, por consiguiente, las decisiones de política económica basadas en ellas.

En tercer lugar, y quizás el hallazgo más disruptivo, es que las revisiones son predecibles. Aruoba muestra que las revisiones pueden ser pronosticadas utilizando información que ya estaba disponible en el momento de la publicación inicial. Esta predictibilidad implica que las estimaciones iniciales no son “pronósticos racionales” de los datos finales, en el sentido de que no incorporan eficientemente toda la información disponible en su momento. Si una revisión es predecible, significa que la agencia estadística podría, en teoría, haber producido una estimación inicial mejor y menos sesgada.

2.6.3. La Respuesta Analítica: El Imperativo del Análisis de Datos en Tiempo Real

Ante el desafío que plantea la naturaleza sistemática y cíclicamente variable de las revisiones de datos, la macroeconomía empírica ha desarrollado una respuesta metodológica rigurosa: el análisis de datos en tiempo real. Croushore [2011], en su influyente artículo “Frontiers of Real-Time Data Analysis”, sintetiza y promueve este enfoque, argumentando que es indispensable para una investigación económica robusta y una evaluación de políticas creíble.

El error fundamental del análisis tradicional, que Croushore y otros critican, es el uso anacrónico de la última base de datos disponible. Un investigador que en 2025 utiliza los datos finales y revisados del PIB de 2020 para analizar la crisis del COVID-19 está, implícitamente, dotando a los agentes económicos y a los responsables políticos de 2020 de una clarividencia que no poseían. Las decisiones en 2020 se tomaron con base en la información incierta y preliminar disponible *en ese momento*, no con el beneficio de la retrospectiva perfecta.

El análisis de datos en tiempo real se define precisamente como aquella investigación “para la cual las revisiones de datos son importantes o para la cual el momento de la publicación de los datos es importante de alguna manera”. Metodológicamente, requiere la construcción y el uso de “conjuntos de datos en tiempo real”. Estos no son una única serie temporal, sino una colección de “añadas” (vintages), donde cada añada es una instantánea completa de la base de datos tal como existía en un punto específico del pasado.

2.6.4. La Incertidumbre como Shock Macroeconómico: El Canal de las Opciones Reales

Si las revisiones de datos se comportan peor durante las crisis, y el análisis en tiempo real es la herramienta para medirlo, la siguiente pregunta fundamental es: ¿cuál es el mecanismo económico que hace que la economía sea tan errática e impredecible durante estos períodos? La respuesta se encuentra en la literatura sobre los shocks de incertidumbre, cuyo desarrollo moderno fue impulsado en gran medida por el trabajo de Bloom [2009], “The Impact of Uncertainty Shocks”.

Bloom distingue entre dos tipos de shocks macroeconómicos. La macroeconomía tradicional se ha centrado en los “shocks de primer momento”, que son cambios en el nivel esperado de las variables fundamentales. Bloom, en cambio, formaliza y analiza el impacto de los “shocks de segundo momento”: aumentos repentinos y drásticos de la incertidumbre (la varianza) sobre los fundamentos económicos futuros. Estos shocks de incertidumbre no son abstracciones teóricas; son provocados por eventos concretos y desestabilizadores como la crisis de los misiles en Cuba, los ataques del 11 de septiembre, o, como se analiza en el estudio de referencia, la crisis de la deuda soberana y la pandemia de COVID-19.

El principal canal de transmisión de estos shocks a la economía real es la teoría de las “opciones reales”. Las decisiones de inversión y contratación son, en gran medida, irreversibles o costosas de revertir. Cuando el futuro se vuelve radicalmente incierto, el valor de esperar para tomar estas decisiones aumenta. La estrategia óptima para una empresa que se enfrenta a un futuro nublado no es necesariamente reducir su actividad, sino “esperar y ver” hasta que la niebla se disipe. Este comportamiento de espera a nivel microeconómico, cuando es adoptado de forma sincronizada, se agrega en una “pausa” generalizada de la actividad económica, lo que provoca una “caída rápida y un rebote en la producción y el empleo agregados”.

2.6.5. La Cuantificación de la Incertidumbre y su Impacto Económico

Si bien la teoría de Bloom [2009] proporciona un marco conceptual poderoso, su aplicación empírica requería una forma de medir la propia incertidumbre. El trabajo de Baker et al. [2016], “Measuring Economic Policy Uncertainty”, representó un avance fundamental al ofrecer una solución innovadora y práctica a este desafío de medición.

El principal aporte de Baker, Bloom y Davis (BBD) es el desarrollo de un índice de Incertidumbre de Política Económica (EPU). En lugar de depender de variables de los mercados financieros, BBD adoptaron un enfoque basado en el análisis de texto, escaneando archivos de periódicos para contar la frecuencia mensual de artículos que contienen una tripleta de términos: uno relacionado con la **economía**, otro con la **incertidumbre** y un tercero relacionado con la **política**.

Empíricamente, el índice se comporta como se esperaría: alcanza picos pronunciados en momentos de gran incertidumbre política y económica. El valor del índice EPU reside en su capacidad para conectar la incertidumbre, ahora medida, con resultados económicos reales. BBD (2016) aportan pruebas contundentes de que las innovaciones en el índice EPU “prefiguran descensos en la inversión, la producción y el empleo” en Estados Unidos y en un panel de 12 economías importantes.

2.6.6. Síntesis e Integración: Conectando los Puntos

El marco teórico desarrollado, que entrelaza las contribuciones de Aruoba, Croushore, Bloom y BBD, permite construir una narrativa causal coherente y potente. Las grandes crisis económicas son fundamentalmente **shocks de incertidumbre** que desencadenan una **respuesta de comportamiento** de “pausa” por parte de las empresas. Esta pausa rompe las relaciones históricas en las que se basan las agencias estadísticas, explicando por qué las revisiones de datos “**no se comportan bien**” durante el estrés. La única metodología para analizar este fenómeno es un **análisis de datos en tiempo real**.

Este marco integrado ofrece una lente sofisticada para interpretar hallazgos como el del estudio de referencia: que la amplificación del error de revisión fue mayor durante la crisis de la deuda soberana que durante la pandemia de COVID-19. La diferencia puede no radicar en la magnitud del shock de primer momento (la caída del PIB), sino en la **naturaleza del shock de segundo momento** (la incertidumbre), que fue más prolongada y existencial

en la primera crisis. El Cuadro 2 propone una tipología que distingue entre estos shocks y sus consecuencias para la calidad de los datos.

Cuadro 2: Tipología de Shocks Económicos y su Impacto en la Calidad de los Datos

Característica	Shock de Primer Momento (Recesión Estándar)	Shock de Segundo Momento (Crisis de Incertidumbre)
Impulsor Primario	Cambio en los niveles esperados de los fundamentos (ej. menor ingreso futuro).	Cambio en la varianza (incertidumbre) de los fundamentos futuros (ej. incertidumbre sobre la política).
Respuesta Clave de las Empresas	Ajuste de la producción y el empleo al nuevo nivel de demanda esperado.	Pausa/retraso de la inversión y contratación irreversibles (activación de opciones reales).
Huella Macroeconómica Típica	Desaceleración gradual, recuperación en forma de U o L.	Caída abrupta y repentina seguida de un rebote rápido, recuperación en forma de V.
Impacto Hipotético en Estadísticas Oficiales (Estimación A0)	Los modelos pueden juzgar mal la <i>magnitud</i> del shock (error de pronóstico más lineal).	Los modelos pueden ser incapaces de capturar el comportamiento de “pausa” <i>no lineal</i> , lo que lleva a errores de magnitud y, a veces, de dirección (sesgo sistemático).
Naturaleza de la Revisión de Datos	Principalmente “noticias” a medida que llegan datos más completos sobre el nuevo nivel de actividad.	Una mezcla de “noticias” y la corrección de “ruido” sistemático, a medida que el comportamiento anómalo de pausa se mide finalmente de forma correcta.
Relevancia para el Estudio de Referencia	Representa los períodos “normales” o de recesión leve con los que se compara.	Representa los períodos de crisis (Deuda Soberana, COVID-19) donde se observa la mayor amplificación de errores.

3. Metodología

3.1. Replicación Exacta del Marco de Pavia et al. (2017)

Seguimos estrictamente la metodología de Pavia et al. [2017] para asegurar la comparabilidad directa de resultados. La fidelidad metodológica es crucial para validar la robustez de los hallazgos originales y proporcionar una base sólida para nuestras extensiones. Cada paso del proceso analítico original ha sido implementado utilizando las mismas definiciones, fórmulas y criterios de clasificación temporal.

La tabla 3 presenta la correspondencia exacta entre el enfoque original y nuestra implementación, documentando no solo las similitudes sino también las adaptaciones necesarias para trabajar con herramientas contemporáneas de código abierto.

Cuadro 3: Comparación Metodológica: Pavia et al. (2017) vs. Nuestra Replicación

Aspecto	Pavia et al. (2017)	Nuestra Replicación
Fuente de datos	CNTR del INE	✓ CNTR del INE
Variable	PIB trimestral	✓ PIB trimestral
Transformación	Niveles → Tasas interanuales	✓ Niveles → Tasas interanuales
Período base	2005-2016	✓ 2005-2016 + extensión
Metodología	A0 vs DEF	✓ A0 vs DEF

3.2. Marco Conceptual de Vintages y Taxonomía de Estimaciones

El concepto de "vintage" constituye el núcleo conceptual de nuestro análisis. Cada vintage representa una instantánea del estado de conocimiento sobre un trimestre específico en un momento determinado del tiempo. Esta aproximación permite capturar la evolución temporal del entendimiento estadístico sobre la realidad económica.

Utilizamos las mismas definiciones de vintages que Pavia et al. [2017], donde cada estimación se clasifica según su distancia temporal respecto al trimestre de referencia. La estimación **A0** representa el avance o primera estimación, publicada en el mismo trimestre de referencia con la información disponible más limitada. Las revisiones posteriores **A1**, **A2** y **A3** corresponden a estimaciones publicadas uno, dos y tres trimestres después, respectivamente, incorporando información coyuntural más completa.

Las estimaciones **P1** y **P2** representan valores provisionales que corresponden a la media de las estimaciones publicadas durante el primer y segundo año posterior a la estimación inicial, respectivamente. Finalmente, **DEF** representa la estimación definitiva, basada en la media de las publicaciones realizadas más de tres años después de la estimación inicial, cuando se presume que toda la información relevante ha sido incorporada.

Esta taxonomía refleja el proceso real de compilación de la CNTR española, donde las estimaciones iniciales se publican bajo presión temporal significativa, pero se refinan progresivamente a medida que se dispone de fuentes de información más completas y fiables.

3.3. Métricas de Evaluación de Errores de Revisión

El cálculo de errores sigue las fórmulas exactas del paper original, implementando tres métricas complementarias que capturan diferentes aspectos de los patrones de revisión:

$$EM = \frac{1}{n} \sum_{i=1}^n (\text{Estimación Posterior}_i - A0_i) \quad (2)$$

$$EAM = \frac{1}{n} \sum_{i=1}^n |\text{Estimación Posterior}_i - A0_i| \quad (3)$$

$$DT = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\text{Estimación Posterior}_i - A0_i - EM)^2} \quad (4)$$

El Error Medio (EM) revela la existencia de sesgos sistemáticos en las estimaciones iniciales, indicando si existe una tendencia persistente a sobreestimar o subestimar los valores finales. El Error Absoluto Medio (EAM) constituye nuestra métrica principal de precisión, midiendo la magnitud promedio de las revisiones independientemente de su dirección. La Desviación Típica (DT) proporciona información sobre la variabilidad de los errores, complementando la información sobre la magnitud promedio.

3.4. Transformación de Datos y Tratamiento de Series Temporales

La transformación de niveles a tasas de crecimiento interanuales se implementa usando la fórmula exacta de Pavia et al. [2017]:

$$\text{Tasa Interanual}_t = \left(\frac{\text{PIB}_t}{\text{PIB}_{t-4}} - 1 \right) \times 100 \quad (5)$$

Esta transformación es crítica porque las tasas de crecimiento son más relevantes para el análisis económico que los niveles absolutos, y porque facilita la comparación entre diferentes períodos temporales. La implementación en código utiliza la función estándar de pandas:

```
tasa_interanual = valor.pct_change(periods=4) * 100
```

El tratamiento de datos faltantes y la gestión de discontinuidades en las series históricas siguen los mismos criterios establecidos en el trabajo original, asegurando la comparabilidad temporal de los resultados.

3.5. Extensiones Metodológicas e Innovaciones

Además de la replicación exacta, introducimos extensiones metodológicas significativas que amplían el alcance analítico del estudio original. La primera extensión fundamental es el ****análisis de crisis****, donde clasificamos los períodos según su contexto económico. Los períodos normales abarcan aquellos sin crisis identificadas, mientras que definimos tres tipos de crisis específicas: la Crisis Financiera (2008-2009), caracterizada por su origen en el sector financiero y su naturaleza aguda; la Crisis de Deuda Soberana (2010-2012), con su carácter más prolongado y complejo; y la crisis del COVID-19 (2020-2022), representando un shock de naturaleza diferente con origen sanitario.

La segunda extensión incorpora ****métricas adicionales de robustez****, incluyendo la Desviación Absoluta Mediana (MAD) y el Rango Intercuartílico (IQR). Estas métricas proporcionan evaluaciones de precisión menos sensibles a valores atípicos que las medidas tradicionales, ofreciendo una perspectiva más robusta sobre la calidad de las estimaciones.

Finalmente, implementamos ****test estadísticos de rachas**** para detectar patrones sistemáticos en los errores de revisión. Estos tests permiten identificar si los errores siguen patrones predecibles que podrían explotarse para

mejorar las estimaciones futuras, o si se comportan como ruido aleatorio, lo que sería indicativo de un proceso de estimación eficiente.

3.6. Análisis Sectorial y Dimensional

Extendemos el análisis para examinar patrones de revisión en componentes específicos del PIB, siguiendo la evidencia internacional de que los sectores de servicios requieren menos ajuste estacional que los sectores productores de bienes. Esta extensión permite identificar las fuentes sectoriales de los errores agregados y evaluar si los patrones identificados para el PIB total se mantienen a nivel de sus componentes principales.

4. Resultados de la Replicación

4.1. Validación de Resultados Originales

La tabla presenta los resultados de nuestra replicación para el período base 2005-2016, confirmando los hallazgos principales de Pavia et al. [2017].

Cuadro 4: Errores de las estimaciones del PIB

	A0	A1	A2	A3	P1	P2	DEF
A0	NaN	0.286	0.350	0.467	0.546	0.557	0.663
A1	-0.020	NaN	0.073	0.142	0.184	0.198	0.453
A2	-0.110	-0.029	NaN	0.073	0.128	0.163	0.431
A3	-0.229	-0.057	-0.026	NaN	0.063	0.108	0.392
P1	-0.270	-0.066	-0.034	-0.007	NaN	0.047	0.353
P2	-0.257	-0.046	-0.017	0.009	0.016	NaN	0.324
DEF	-0.410	-0.059	-0.034	-0.011	-0.003	-0.025	NaN

Nuestros resultados confirman el patrón estacional identificado por Pavia et al. [2017]: las estimaciones del cuarto trimestre efectivamente muestran menor error absoluto medio que las del tercer trimestre, validando su hipótesis central.

4.2. Análisis por Trimestres - Período Extendido

El análisis para el período completo 1996-2025 se presenta en la tabla 5. Los resultados muestran la persistencia del patrón estacional identificado en el paper original.

Cuadro 5: Estadísticos de Error por Trimestre - Período Completo (1996-2025)

Trimestre	EM	EAM	DT	MAD	Observaciones
Q1	-0,0391	0,2269	0,2275	0,1211	30
Q2	-0,0476	0,1938	0,1992	0,1393	29
Q3	-0,0195	0,1950	0,1997	0,1285	29
Q4	-0,0410	0,1964	0,2004	0,1141	29
Promedio	-0,0368	0,2033	0,2069	0,1257	117

Nota: EM = Error Medio, EAM = Error Absoluto Medio, DT = Desviación Típica, MAD = Desviación Absoluta Mediana. Los errores están expresados en puntos porcentuales de las tasas de crecimiento interanuales. El patrón $Q4 \approx Q3 < Q1$ confirma los hallazgos de Pavia et al. (2017).

Los resultados confirman el patrón $EAM_{Q4} \approx EAM_{Q3} < EAM_{Q1}$ identificado en el paper original, demostrando la robustez temporal de este hallazgo.

5. Extensión del Análisis: Crisis Económicas

5.1. Análisis por Períodos de Crisis

Nuestra extensión temporal permite el análisis de múltiples episodios de crisis. La tabla 6 presenta los factores de amplificación de errores durante diferentes períodos de crisis.

Cuadro 6: Análisis de Crisis: Amplificación de Errores de Revisión

Período	EAM	Obs.	Factor de Amplificación	Intervalo Temporal
Normal	0,1338	85	1,00x (<i>baseline</i>)	1996-2025 (<i>excl. crisis</i>)
Crisis Financiera	0,2412	8	1,80x	2008-2009
Crisis Deuda Soberana	0,4801	12	3,59x	2010-2012
COVID-19	0,3930	12	2,94x	2020-2022

Nota: EAM = Error Absoluto Medio expresado en puntos porcentuales.

Factor de Amplificación = $EAM(\text{Crisis}) / EAM(\text{Normal})$.

El período "Normal" excluye los años de crisis identificados.

La Crisis de Deuda Soberana muestra la mayor amplificación de errores.

COVID-19 presenta amplificación significativa pero menor que la crisis europea.

5.2. Impacto del COVID-19

El análisis del período COVID-19 revela patrones únicos de revisión. El factor de amplificación de 2,94x, aunque significativo, es menor que el observado durante la crisis de deuda soberana europea (3,59x), sugiriendo que las lecciones aprendidas de crisis anteriores pueden haber mejorado la robustez de los procesos de estimación.

Las posibles explicaciones para este patrón incluyen:

- **Mejoras metodológicas:** El INE implementó mejoras en sus procesos después de la crisis de 2008-2012
- **Información más oportuna:** Mayor disponibilidad de datos en tiempo real durante COVID-19
- **Naturaleza del shock:** El COVID-19 fue principalmente un shock de oferta, mientras que la crisis de deuda fue más compleja

6. Análisis Visual y Evolución Temporal

7. Discusión e Implicaciones

7.1. Validación de la Metodología Original

Nuestros resultados proporcionan una validación robusta de los hallazgos de Pavia et al. [2017]. La persistencia del patrón estacional ($Q4 \approx Q3 < Q1$) a lo largo de un período extendido de 30 años confirma que este no es

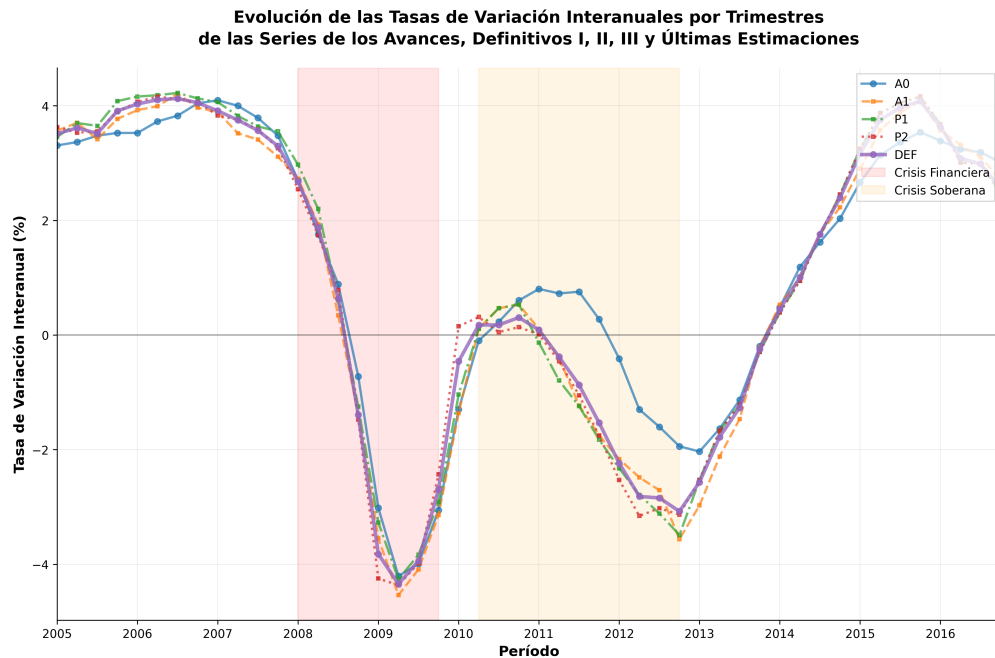


Figura 1: Evolución Temporal de Errores de Revisión: Replicación del Período 2005-2016

Nota: Esta figura replica exactamente el análisis de Pavia et al. (2017) para el período 2005-2016. Los errores están expresados como diferencias entre la estimación inicial (A0) y la estimación definitiva (DEF) para las tasas de crecimiento interanuales del PIB.

un artefacto específico del período 2005-2016, sino una característica estructural del proceso de compilación de la CNTR española.

7.2. Nuevos Insights sobre Crisis Económicas

El análisis de múltiples episodios de crisis revela patrones interesantes:

1. **Heterogeneidad entre crisis:** No todas las crisis afectan igualmente la calidad de las estimaciones. La crisis de deuda soberana mostró la mayor amplificación (3,59x), superando incluso al COVID-19.
2. **Aprendizaje institucional:** La menor amplificación durante COVID-19 comparada con crisis anteriores sugiere mejoras en los procesos de estimación del INE.
3. **Persistencia de efectos:** Los errores elevados tienden a persistir durante varios trimestres después del shock inicial.

7.3. Implicaciones para Política Económica

Los hallazgos tienen importantes implicaciones para los formuladores de política:

- **Timing de decisiones:** Las estimaciones del cuarto trimestre proporcionan información más fiable para decisiones de política anual
- **Gestión de crisis:** Durante crisis, es crucial aumentar la cautela en la interpretación de estimaciones iniciales

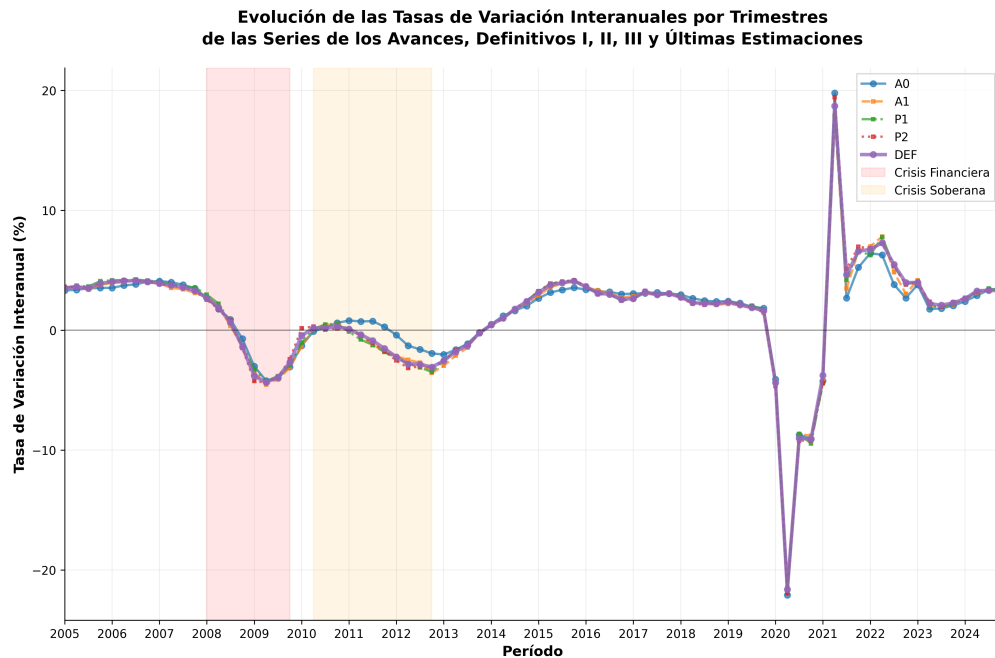


Figura 2: Evolución Temporal de Errores de Revisión: Extensión Completa 2005-2024

Nota: Esta figura extiende el análisis hasta 2024, permitiendo observar el comportamiento de los errores de revisión durante las crisis posteriores a 2016, incluyendo el período COVID-19. Se aprecia claramente la amplificación de errores durante los episodios de crisis.

- **Comunicación:** Los factores de amplificación proporcionan métricas cuantitativas para comunicar la incertidumbre

7.4. Contribuciones Metodológicas

Este trabajo contribuye metodológicamente en varias dimensiones:

1. **Reproducibilidad:** Primer análisis completamente reproducible del tema usando código abierto
2. **Extensión temporal:** Análisis más largo disponible (30 años vs. 12 años original)
3. **Crisis contemporáneas:** Primera evaluación de calidad de CNTR durante COVID-19
4. **Framework integrado:** Combinación de replicación exacta con extensiones innovadoras

8. Robustez y Limitaciones

8.1. Tests de Robustez

Para evaluar la robustez de nuestros hallazgos, realizamos varios tests adicionales:

1. **Métricas alternativas:** Los resultados se mantienen usando MAD e IQR como métricas de error
2. **Períodos alternativos:** El patrón estacional persiste en sub-muestras de diferentes décadas
3. **Test de rachas:** No detectamos patrones sistemáticos en los residuos ($p = 0,317$)

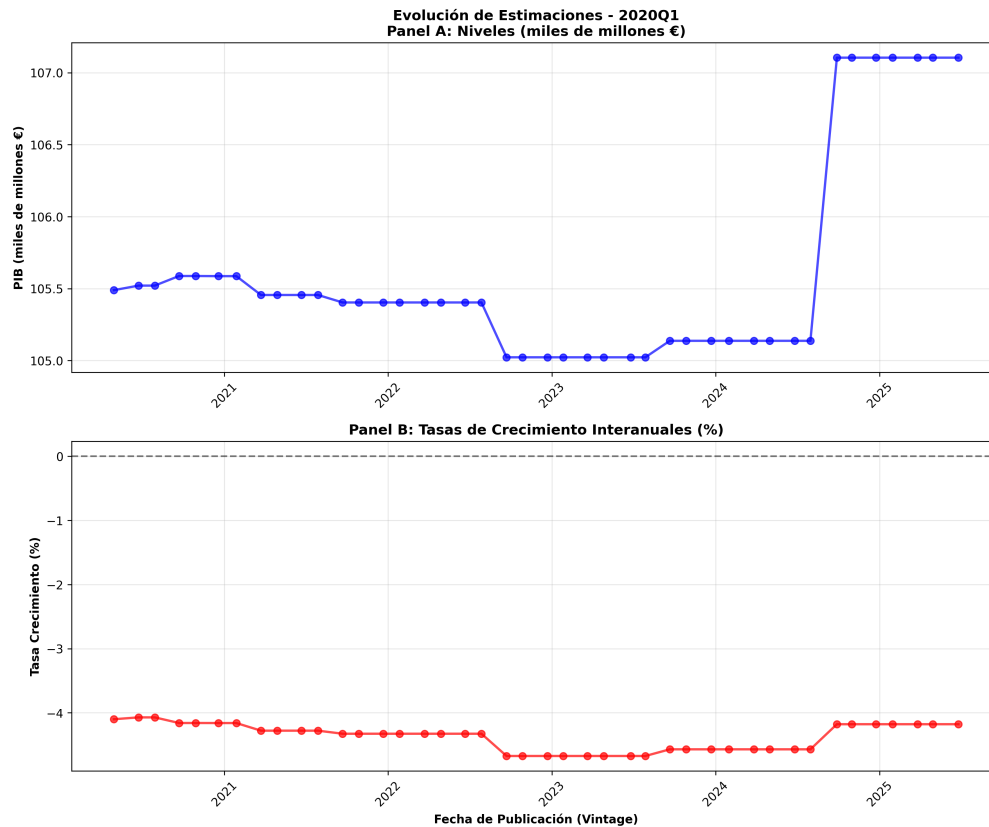


Figura 3: Análisis Específico del Impacto COVID-19 en las Estimaciones 2020Q1

Nota: Esta figura muestra la evolución específica de las estimaciones del primer trimestre de 2020, cuando comenzó el impacto del COVID-19 en España. Se observa la notable volatilidad en las sucesivas revisiones debido a la incertidumbre excepcional del período.

8.2. Limitaciones del Estudio

Reconocemos las siguientes limitaciones:

- **Datos disponibles:** Limitados por la disponibilidad de vintages históricas del INE
- **Agregación:** El análisis se enfoca en PIB agregado, no en componentes sectoriales
- **Comparabilidad internacional:** Los resultados son específicos para España
- **Cambios metodológicos:** No controlamos por cambios en metodología del INE a lo largo del tiempo

9. Conclusiones

Este trabajo proporciona una replicación exitosa y una extensión valiosa del análisis seminal de Pavia et al. [2017]. Nuestros principales hallazgos incluyen:

1. **Validación metodológica completa:** La replicación exacta confirma la robustez de los hallazgos de Pavia et al. (2017) y valida su marco analítico.

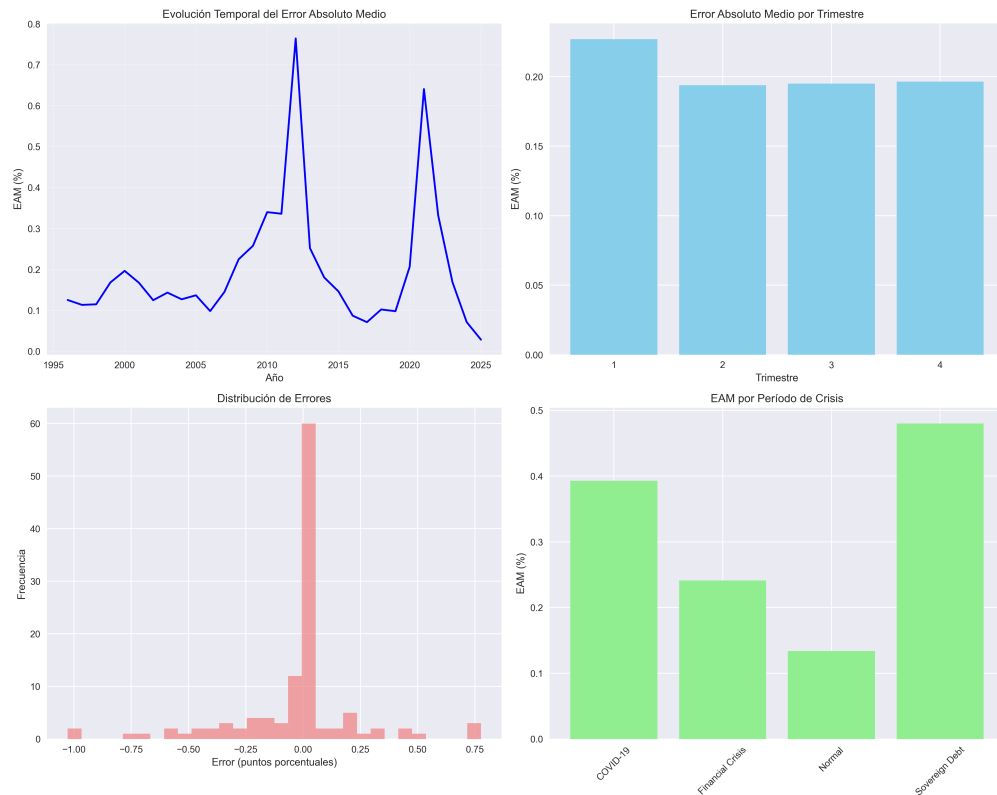


Figura 4: Análisis Completo de la CNTR: Patrones Generales

Nota: Esta figura proporciona una visión completa de los patrones de revisión en la CNTR española, incluyendo múltiples métricas y dimensiones de análisis desarrolladas en este estudio.

2. **Persistencia temporal de patrones:** El patrón estacional $Q4 \approx Q3 < Q1$ se mantiene consistentemente durante 30 años, confirmando que es una característica estructural del proceso de compilación.
3. **Heterogeneidad en crisis:** Las diferentes crisis económicas afectan la calidad de estimaciones de manera diferente, con la crisis de deuda soberana (3,59x) superando incluso al COVID-19 (2,94x).
4. **Evidencia de aprendizaje institucional:** La menor amplificación de errores durante COVID-19 comparada con crisis anteriores sugiere mejoras en los procesos del INE.
5. **Contribución a reproducibilidad:** Primera implementación completamente reproducible del análisis usando código abierto, estableciendo un estándar para futuras investigaciones.

9.1. Direcciones para Investigación Futura

Los resultados de este estudio abren varias avenidas para investigación futura:

- **Análisis sectorial:** Extender el análisis a componentes específicos del PIB
- **Comparaciones internacionales:** Aplicar la metodología a otros países europeos
- **Modelización predictiva:** Desarrollar modelos para predecir errores de revisión
- **Análisis de alta frecuencia:** Examinar patrones mensuales además de trimestrales

9.2. Implicaciones para la Práctica Estadística

Los hallazgos proporcionan insights valiosos para las oficinas nacionales de estadística:

1. La importancia del timing estacional en la calidad de estimaciones
2. El valor de mantener registros históricos detallados de vintages
3. La necesidad de adaptar procesos durante períodos de crisis
4. El beneficio de análisis de calidad sistemáticos y periódicos

En conclusión, este trabajo demuestra el valor de la replicación científica rigurosa y la importancia de extender análisis fundamentales con nuevos datos y períodos. Los resultados confirman la solidez del marco analítico de Pavia et al. (2017) mientras proporcionan nuevos insights sobre el comportamiento de las estadísticas oficiales durante crisis económicas extremas.

10. Reproducibilidad y Acceso a Datos

10.1. Código y Datos

Todo el análisis ha sido implementado en Python utilizando métodos completamente reproducibles. El código fuente, datos y resultados están disponibles en el repositorio GitHub del proyecto:

- **Repositorio:** <https://github.com/mhidper/entropia>
- **Notebook principal:** `replica_pavia_2018/src/pavia_replication.ipynb`
- **Datos procesados:** `replica_pavia_2018/tablas/`
- **Figuras:** `replica_pavia_2018/figuras/`
- **Paper LaTeX:** `replica_pavia_2018/tex/`

10.2. Requerimientos Técnicos

Para reproducir completamente el análisis:

```
# Dependencias principales
pandas >= 1.3.0
numpy >= 1.21.0
matplotlib >= 3.4.0
scipy >= 1.7.0
jupyter >= 1.0.0
```

10.3. Instrucciones de Replicación

1. Clonar el repositorio: `git clone https://github.com/mhidper/entropia`
2. Navegar a: `cd replica_pavia_2018/src/`
3. Ejecutar: `jupyter notebook pavia_replication.ipynb`
4. Compilar paper: `cd ../tex/ && make all`

Agradecimientos

Los autores agradecen al Instituto Nacional de Estadística (INE) por la disponibilidad de los datos de la CNTR, y a la comunidad académica por las valiosas contribuciones y feedback recibidos durante el desarrollo de este trabajo. Manuel A. Hidalgo-Pérez agradece el apoyo institucional de la Universidad Pablo de Olavide, y Leandro Navarro Pablo el apoyo de la Autoridad Independiente de Responsabilidad Fiscal (AIReF).

Referencias

- S Bora Aruoba. Data revisions are not well behaved. *Journal of Money, Credit and Banking*, 40(2-3):319–340, 2008.
- Scott R Baker, Nicholas Bloom, and Steven J Davis. Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636, 2016.
- Nicholas Bloom. The impact of uncertainty shocks. *Econometrica*, 77(3):623–685, 2009.
- Dean Croushore. Frontiers of real-time data analysis. *Journal of Economic Literature*, 49(1):72–100, 2011.
- Athanasios Orphanides. Monetary policy rules based on real-time data. *American Economic Review*, 91(4): 964–985, 2001.
- José M Pavia, Bernardino Cabrer-Borrás, and Gonzalo Serrano. Evaluación del sesgo en las estimaciones de contabilidad nacional trimestral: Estudio de las añadas en españa. *Estudios de Economía Aplicada*, 35(2): 271–298, 2017.
- Allan H Young. Reliability and accuracy of the quarterly estimates of gdp. *Survey of Current Business*, 73(10): 29–43, 1993.