UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

TESIS DOCTORAL

Four essays on the application of nonlinear techniques to time series in economics

Cuatro ensayos sobre la aplicación de técnicas no lineales a series temporales en economía

D. Salvador Ramallo Ros

2023

UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

TESIS DOCTORAL

Four essays on the application of nonlinear techniques to time series in economics

Cuatro ensayos sobre la aplicación de técnicas no lineales a series temporales en economía

D. Salvador Ramallo Ros

2023

Dirección:
Prof. Dr. D. Máximo Cosme Camacho Alonso

UNIVERSIDAD DE
MURCIA

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA EN MODALIDAD DE COMPENDIO O ARTÍCULOS PARA OBTENER EL TÍTULO DE DOCTOR

Aprobado por la Comisión General de Doctorado el 19-10-2022

D./Dña. Salvador Ramallo Ros

doctorando del Programa de Doctorado en

Interuniversitario en Economia (DECiDE)

de la Escuela Internacional de Doctorado de la Universidad Murcia, como autor/a de la tesis presentada para la obtención del título de Doctor y titulada:

Four Essays on the Application of Nonlinear Techniques to Time Series in Economics

y dirigida por,

D./Dña. Maximo Cosme Camacho Alonso

D./Dña.

D./Dña.

DECLARO QUE:

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita, cuando se han utilizado sus resultados o publicaciones.

Además, al haber sido autorizada como compendio de publicaciones o, tal y como prevé el artículo 29.8 del reglamento, cuenta con:

- *La aceptación por escrito de los coautores de las publicaciones de que el doctorando las presente como parte de la tesis.*
- *En su caso, la renuncia por escrito de los coautores no doctores de dichos trabajos a presentarlos como parte de otras tesis doctorales en la Universidad de Murcia o en cualquier otra universidad.*

Del mismo modo, asumo ante la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada, en caso de plagio, de conformidad con el ordenamiento jurídico vigente.

En Murcia, a 13 de Octubre de 2023

Fdo.: Salvador Ramallo Ros

Firmado por RAMALLO ROS SALVADOR - ***9205** el día 13/10/2023 con un certificado emitido por AC FNMT Usuarios

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe ser insertada en la primera página de la tesis presentada para la obtención del título de Doctor.

Información básica sobre protección de sus datos personales aportados	
Responsable:	Universidad de Murcia. Avenida teniente Flomesta, 5. Edificio de la Convalecencia. 30003; Murcia. Delegado de Protección de Datos: dpd@um.es
Legitimación:	La Universidad de Murcia se encuentra legitimada para el tratamiento de sus datos por ser necesario para el cumplimiento de una obligación legal aplicable al responsable del tratamiento. art. 6.1.c) del Reglamento General de Protección de Datos
Finalidad:	Gestionar su declaración de autoría y originalidad
Destinatarios:	No se prevén comunicaciones de datos
Derechos:	Los interesados pueden ejercer sus derechos de acceso, rectificación, cancelación, oposición, limitación del tratamiento, olvido y portabilidad a través del procedimiento establecido a tal efecto en el Registro Electrónico o mediante la presentación de la correspondiente solicitud en las Oficinas de Asistencia en Materia de Registro de la Universidad de Murcia

La sabiduría no es producto de la educación, sino del intento de toda la vida para adquirirla.
Albert Einstein

A mis padres

Agradecimientos

Al enfrentarme a este folio como últimas lineas a redactar de este proyecto, me doy cuenta de la dificultad que implica resumir en unas líneas el gran número de personas a las que estoy agradecido.

En primer lugar, Máximo. Soy consciente de lo afortunado que soy al tener la oportunidad de trabajar contigo. Agradezco tu infinita paciencia, tu constante interés y tu capacidad para transmitirme que investigar divirtiéndose es la mejor forma de hacerlo. Sin ti, estoy seguro de que seguiría a día de hoy añadiendo anexos al primer trabajo. No sólo me has apoyado en este proyecto, si no que también me has ayudado a comprender y adoptar una filosofía de trabajo. Pocos pueden decir que su director de tesis es también uno de sus mejores amigos. Y este es solo el primero de muchos proyectos.

En segundo lugar, a tí Manolo. Siempre has estado atento sin tener ningún motivo para hacerlo. Desde el principio he disfrutado trabajando contigo y te he sentido al lado de Máximo en este viaje. Al igual que con Máximo, he tenido mucha suerte de que el atractor de mi dinámica caótica fueseis vosotros (y una vez converge a él, de ahí no se separa esa dinámica). También quiero agradecer a Andrés por compartir muchos ratos conmigo, ya sea disfrutando de la belleza de la ciencia o de la que también hay en una conversación de actualidad.

Debo de destacar también la importancia en esta etapa formativa de los proyectos del ministerio PID2019-107800GB-I00 y PID2022-136547NB-I00, al Grupo de Excelencia Economic Modelling and Non Parametric Statistics de la Fundación Séneca. También de la Fundación Fulbright, tanto por su financiación para poder realizar una estancia de un valor incalculable como por ser una Fundación con unos valores admirables. También agradecer al profesor Porfiri por aceptarme en su laboratorio durante ese período y por la oportunidad de aprender de su brillantez y de su más que constante dedicación. Por supuesto, agradezco a los compañeros del lab, especialmente a Alain por aprender de tu excelencia y humildad, a Anna por su disposición constante, y sobre todo a Rayan, un amigo con todas las letras con el que compartiré muchos proyectos y momentos en el futuro.

También quiero agradecer a los distintos investigadores anónimos de los congresos a los que pude asistir gracias a los mencionados proyectos. En general, escuchar a grandes investigadores a los que tanto respetas, aprender de ellos y recibir su feedback es un lujo. El gesto de aprobación o una sonrisa de personas de tan alto nivel ayuda a cualquier joven e inseguro investigador.

Con un carácter especial, quiero agradecer a mis padres por su apoyo y su trabajo constante, inspirador desde que tengo uso de razón. Sin vosotros no estaría tan seguro de que sin un trabajo duro y constante las cosas no llegan. Mucha culpa de este trabajo es vuestra. También quiero agradecer a mi hermana, siempre atenta y protectora como buena hermana mayor, y una trabajadora incansable y fuerte como pocas personas. Otro espejo en el que mirarme. Y por supuesto, tengo que agradecerle a Lorena por su eterna paciencia. Has estado a mi lado en todo este camino, sufriendo los muchos días peores y los pocos días mejores. Sin ti, esto seguro no hubiese llegado a buen puerto. Muchas gracias por estar siempre y aportarme ese equilibrio

necesario.

A riesgo de dejarme a muchas personas que han aportado su granito de arena en el plano personal, quiero agradecer a demás familia (tíos y tías, abuelos y abuelas, primos y primas), a Mercedes, a mis amigos y amigas de uno y otro sitio (vecinos, mazarroneros, de Inglaterra, del instituto, universidad, sardineros hercúleos, etc), a los que no voy a particularizar para no dejarme a ninguno. A esos compañeros del BBVA con los que tanto aprendí, que fueron pacientes conmigo y que les tengo un cariño especial. A mis compañeros y compañeras de la UNIR, por ser tan fácil trabajar con ellos y por darme una oportunidad en su momento. También a los compañeros de doctorado, siempre necesarios para dar ánimos, al doctorado DECIDE y a Fernando como su coordinador, y por supuesto al departamento de Métodos Cuantitativos para la Economía y Empresa de la Universidad de Murcia, en particular a Arielle que fue quién me puso en la puerta de la persona adecuada. También debo agradecer a Samuel por estar disponible con una sonrisa siempre que hacía falta, y a Adolfo por sus consejos y sabiduría.

Resumen

Modelización no lineal

El método principal y más utilizado para dilucidar las relaciones, las dinámicas y, en general, para modelizar el comportamiento tanto de los individuos como de los indicadores macroeconómicos, ha sido de forma habitual el enfoque lineal. Su sencillez permite a los economistas comunicar las ideas y relaciones económicas de forma clara y concisa, lo que facilita su interpretación (un aspecto esencial de la toma de decisiones económicas y políticas). Además, permite un análisis intuitivo, por ejemplo, de elasticidades y previsiones.

Sin embargo, rara vez la realidad se rige por un comportamiento lineal. La economía es un sistema complejo que comprende múltiples variables interdependientes que pueden interactuar de forma no lineal, dando lugar a resultados imprevisibles o contraintuitivos cuando se emplean métodos lineales tradicionales.

Existen numerosos casos en los que la macroeconomía o la microeconomía no se ajustan a un comportamiento lineal. Por ejemplo, las respuestas a las perturbaciones pueden describirse utilizando funciones de respuesta al impulso no lineales en el caso de series temporales persistentes, como se ve en Potter (1998) en relación con el PNB de EE.UU., mientras que, por su parte, el crédito puede actuar como propagador no lineal de las perturbaciones económicas (Balke, 2000). Además, los cambios en las variables económicas pueden dar lugar a respuestas no proporcionales en otras variables, como ocurre con el consumo en respuesta a perturbaciones de la riqueza, la renta y los tipos de interés (Coskun, Apergis y Coskun, 2022), o en cómo los mercados bursátiles (Escobari y Sharma, 2020) o el PIB (Karaki, 2017) responden a las perturbaciones del precio del petróleo.

La mayor evidencia y una creciente disponibilidad de datos han supuesto que la aplicación de técnicas no lineales en economía haya cobrado importancia en los últimos años, pudiendo éstas aportar múltiples ventajas. La primera de ellas es la mayor precisión de los modelos y las previsiones. La dinámica económica presenta con frecuencia dinámicas que no son sencillas, como efectos umbral, caos o efectos de segundo orden. Las técnicas lineales, incapaces de acomodar estas complejidades, se quedarán cortas en términos de precisión. Por lo tanto, los modelos no lineales, menos propensos a sufrir en esos casos, pueden conducir a una mayor precisión de las previsiones.

Aunque la modelización no lineal se remonta al menos a Hicks (1950) y Goodwin (1951) como solución para abordar el problema de la trayectoria temporal oscilatoria que aparecía en el enfoque lineal como, por ejemplo, en el enfoque de Samuelson (1939), fue después del desarrollo de los modelos lineales en series temporales con el influyente artículo de Box y Jenkins (1976) que popularizó los modelos ARIMA (basado en el trabajo seminal de Whittle en 1951), y su extensión al contexto multivariante por Sims (1980), cuando más se desarrollan los modelos no lineales buscando ampliar estos planteamientos.

Debido a que los modelos ARIMA y multivariantes antes mencionados se basan en el supuesto de linealidad y simetría, que rara vez se cumplen en datos económicos, Tong (1978) introdujo el concepto de efectos umbral en el modelo autorregresivo de umbral (TAR por sus siglas en inglés) para tener en cuenta la asimetría, incorporando más tarde Anderson y Teräsvirta (1992) una transición suave en el umbral mediante el modelo autorregresivo de transición suave (STAR). Para los casos en los que el cambio de régimen se asume modelizable, Tong (1990) propuso los modelos autorregresivos de umbral de autoexcitación (SETAR) en los que las observaciones retardadas de la variable dependiente provocan cambios de régimen, mientras que Hamilton (1989) supone un proceso de Markov como base de estos cambios. Estos modelos se han extendido a un contexto multivariante, como el modelo autorregresivo vectorial de umbral (TVAR) (Tsay, 1998; Hubrich y Teräsvirta, 2013), los modelos autorregresivos vectoriales de transición suave (Camacho, 2004; Teräsvirta y Yang, 2014), el modelo multivariante con umbral de autoexcitación (MSETAR) de Arnold y Günther (2001), y los modelos de Vectores autorregresivos Markov Switching (MSVAR) en Krolzig y Krolzig (1997).

No obstante, la introducción de estas dinámicas más sofisticadas requiere a veces de técnicas no paramétricas tanto para la estimación como para la elaboración de tests, técnicas éstas menos propensas a sufrir cuando los datos presentan no linealidades. Las técnicas de Boostrap, por ejemplo, se han aplicado para probar los efectos de umbral en Giannerini, Goracci y Rahbek (2021), y como técnica de estimación en el modelo de Markov Switching en Ho (2001), mientras que el algoritmo genético se ha empleado como técnica de estimación en los modelos MSETAR en Baragona y Cucina (2013). En el mismo contexto de extensión de modelizaciones lineales, cabe mencionar el modelo EGARCH (Nelson, 1991), el GJR (Glosten, Jagannathan y Runkle, 1993), o el GARCH de transición suave (Hagerud, 1997; Lanne y Saikkonen, 2005) como extensiones no lineales de los modelos GARCH (Bollerslev, 1986), los cuales por si sólos no consiguen captar correctamente la volatilidad asimétrica, como señalan Bildirici y Orsen (2014). Además también se han propuesto combinaciones híbridas, como por ejemplo el modelo Fuzzy-EGARCH-ANN (Mohammed, Aduda y Kubo, 2020) para captar mejor las asimetrías en la volatilidad de los rendimientos financieros. Por último, los modelos de factores dinámicos (Stock y Watson, 1991), ampliamente utilizados para la previsión del ciclo económico, incorporan en la extensión de Brockwell y Davis (2009) la no linealidad manteniendo el filtro de Kalman sin actualizar cuando se manejan valores ausentes, aumentando la propuesta inicial a un marco no lineal. Otra forma de extensión no lineal dentro de estos modelos es la forma de dinámica de segundo orden propuesta en Guerrón-Quintana Khazanov, y Zhong (2021) para la elaboración de un índice de actividad económica y que se estima a través del filtro de Kalman sin esencia (UKF por sus siglas en ingles).

Una segunda característica a destacar cuando se profundiza en la modelización no lineal de series temporales es su mayor capacidad para *manejar bases de datos extensas con variables interrelacionadas*. En los últimos años, los datos económicos han aumentado en tamaño y complejidad, dando lugar a conjuntos de datos más grandes con relaciones entrelazadas. Ciertas técnicas no lineales, como los modelos basados en agentes y las técnicas de aprendizaje automático, resultan útiles para manejar datos tan voluminosos. Es más, las técnicas no lineales podrían ayudar a descubrir relaciones que no se detectarían con técnicas lineales. Aunque algunas de estas técnicas se remontan en el tiempo, su utilidad se ha hecho con el tiempo cada vez más evidente debido a la proliferación de la capacidad de almacenamiento de datos, lo que ha dado lugar a una adopción más amplia en las ciencias sociales y la economía. Por ejemplo, las técnicas de Random Forest (RF) (Breiman, 2001) se han aplicado recientemente a la predicción del crecimiento del PIB y la pobreza (Adriansson y Matterson, 2015; Sohnesen, 2017) empleando herramientas como las medidas de importancia de las variables y los gráficos

de dependencia parcial para explorar las relaciones económicas. El Gradient Boosting (Friedman, 2001) se ha utilizado para predecir quiebras en el sector bancario de EE.UU. (Carmona, Climent y Momparler, 2019) y para predecir el crecimiento del PIB real (Yoon, 2021). Las máquinas de vectores de soporte (Boser et al., 1992) se han utilizado para predecir los precios de los mercados energéticos en Papadimitrou, Gogas y Stathakis (2014), mientras que las redes neuronales recurrentes (Rumelhart et al. 1986) se han aplicado, entre otros, para predecir los tipos de cambio en Kuan y Liu (1995), y para predecir la inflación no lineal en Almosova y Andresen (2023).

De hecho cuando la economía, y en particular la econometría, surge como disciplina, la escasez de datos transversales y longitudinales hizo que la modelización lineal fuera más atractiva que la no lineal. Los enfoques no lineales se consideraban menos adecuados para extraer información ante la escasez de datos. Sin embargo, en la actualidad, la capacidad de recopilación de datos, la disponibilidad de numerosos datos históricos y la mejora de las capacidades informáticas han hecho que los modelos no lineales sean más atractivos y asequibles para este contexto.

Además, y en relación con la mayor disponibilidad de datos mencionada y el deseo de ampliar la modelización lineal, en los últimos años ha aumentado la bibliografía sobre datos de panel no lineales. La creciente capacidad de almacenamiento ha hecho que los datos de series temporales de varios individuos sean fácilmente accesibles, lo que hace que los datos de panel sean especialmente relevantes. Para incorporar las no linealidades a la modelización de datos de panel, Hansen (1999) desarrolló un modelo de umbral de panel, posteriormente ampliado a un marco de panel de transición suave por González, Teräsvirta y Van Dijk (2004). Por su parte, Kremer et al. (2013) proponen un modelo de panel dinámico de umbral, estimado mediante GMM, que permite la presencia de regresores endógenos pero que exige exogeneidad a la variable de transición. Seo y Shin (2016) generalizaron este enfoque permitiendo regresores endógenos y exógenos también en la variable de transición.

Por lo tanto, las técnicas no lineales en el análisis de series temporales ofrecen la posibilidad de modelizar la complejidad de los sistemas económicos, de proporcionar descripciones sobre las dinámicas más ricas y precisas, y mejorar la exactitud de las predicciones, lo que, en última instancia, conduce a una toma de decisiones más informada y rigurosa.

Contribución

Esta tesis pretende contribuir a esta literatura sobre modelización de series temporales no lineales mediante su aplicación en una serie de distintos análisis. Para ofrecer una visión general de estas contribuciones, junto con la estructura de la tesis, podemos resumirlas como sigue. En el capítulo 2, proponemos un modelo univariante no paramétrico para predecir las recesiones fechadas por el Economic Cycle Research Insitute (ECRI) de los países del G7. Se trata de una cuestión relevante, sobre todo en tras las profundas repercusiones económicas de la pandemia de Covid-19, donde muchos modelos de previsión han perdido precisión. En el Capítulo 3, utilizamos un modelo de arboles de decisión basado en boosting para examinar qué indicadores económicos predicen mejor las recesiones en la economía española para una muestra de 270 indicadores mensuales a lo largo de 50 años. La disponibilidad de bases de datos extensas y exhaustivas ha hecho cada vez más factible la utilización de técnicas capaces de identificar relaciones no inmediatas y multidimensionales. En el capítulo 4, proponemos la aplicación de un modelo factorial dinámico a los homicidios con armas de fuego en EE.UU.,

donde el modelo se estima mediante un filtro de Kalman no lineal debido a la presencia de observaciones ausentes, y compara su rendimiento con otros modelos lineales y no lineales. El retraso en la publicación de los datos oficiales en esta serie tiene implicaciones significativas para los responsables políticos y agentes económicos como aseguradoras o inmobiliarios, provocando que ésta investigación sea de especial relevancia. En el Capítulo 5, proponemos un modelo de panel dinámico no lineal, basado en un modelo de gravedad, para explicar los efectos spillover (o de influencia cruzada) entre países de la zona del euro, previamente estimados mediante la metodología de descomposición de la varianza de los errores de predicción de un modelo VAR. Entender los distintos grados de interconexión entre países a lo largo del tiempo es de suma importancia y tiene implicaciones para la política económica y monetaria.

Además de sus aportaciones sustantivas, cada uno de estos capítulos también ofrece valor en términos de aportación de código. Algunos capítulos presentan nuevas propuestas de modelos, mientras que otros implican la adaptación de modelos de lenguajes de programación como Gauss o Matlab a R. Esta dimensión transversal añade profundidad y practicidad a la investigación, haciéndola no sólo académicamente significativa, sino también accesible y aplicable en contextos del mundo real.

Modelización univariante no lineal no paramétrica para la predicción de recesiones económicas

Los sistemas de alerta temprana desempeñan un papel clave en la capacidad de los agentes económicos para tomar decisiones de inversión y ahorro adecuadas. También influyen significativamente en las decisiones políticas de los gobiernos y los responsables políticos. En consecuencia, en la literatura académica existen numerosas propuestas de sistemas de alerta temprana: desde el famoso artículo de Hamilton que propone un modelo de Markov Switching (1989), hasta el enfoque STAR defendido por Anderson y Terasvirta (1992), pasando por el modelo modelo probit de Estrella y Mishkin (1998). Un poco antes, Wecker (1979) hizo una propuesta de previsión basada en un modelo autorregresivo, que posteriormente Camacho y Pérez Quirós (2002) confirmaron que tenía una considerable capacidad empírica. Para este tipo de análisis se suelen comparar los periodos identificados como de alta probabilidad de recesión por el modelo con los periodos identificados por el el National Bureau of Economic Research (NBER) y el ECRI.

No obstante, un acontecimiento sin precedentes desde que se dispone de datos económicos como ha sido el colapso económico causado por la pandemia COVID-19 y las rápidas medidas anticíclicas aplicadas por los responsables políticos, provocó una caída de la actividad económica de una magnitud nunca antes registrada seguida de un repunte de magnitud similar, lo que ha causado importantes problemas de ajuste en los modelos antes mencionados. Algunas propuestas tratan esta recesión de forma única y especial para adaptar enfoques paramétricos (Leiva-León, Perez-Quirós y Rots, 2020; Carreiro et al., 2021; Ng, 2021), mientras que los indicadores de instituciones como el Econbrowser (índice de indicadores de recesión basado en el PIB en base al modelo de Chauvet y Hamilton (2006)) fija los parámetros para el segundo trimestre de 2020 y no actualiza su cálculo en ese periodo, mientras que por su parte el indicador de probabilidad de recesión de la Reserva Federal de San Luis (basado en el modelo de Chauvet y Piger, 2008), estima permitendo un cambio de los parámetros del modelo en dicho trimestre. Alternativamente, McGrane (2022) deja de estimar los parámetros a finales de 2019 como solución.

El primer capítulo de esta tesis contribuye a esta literatura proponiendo un enfoque no lineal y no paramétrico para la previsión de recesiones, que no requiere modificar la base de datos ni

el modelo, y que es robusto frente a observaciones influyentes. La propuesta se construye sobre la base de la metodología de Wecker. Esta ampliación permite generar de previsiones para un período de tiempo futuro (h-períodos hacia delante) en un período de tiempo t dado mediante el aumento de la serie temporal actual con incrementos históricos encontrados en bloques de tamaño $(h + 1)$ extraídos del pasado de la serie temporal. Estas trayectorias pronosticadas se utilizan posteriormente para calcular la frecuencia relativa de las recesiones técnicas con ponderación de las distintas trayectorias de predicción posibles, obtenidas incrustando la serie en el espacio de símbolos. Mediante una simulación de Monte Carlo, analizamos la capacidad de predicción de la propuesta frente a la aproximación de Wecker y frente a un modelo de Markov Switching, obteniendo un rendimiento robusto en nuestro caso frente a los problemas de la competencia. Se realiza Un ejercicio empírico para los países del G7, obteniendo una mayor capacidad predictiva en el ejercicio dentro de muestra (in-sample) tras incorporar la observación derivada de la pandemia.

Si bien el primer capítulo centra su contribución al uso de modelos no lineales y no paramétricos en series temporales univariantes con procesos generadores de datos que presentan no linealidades, se trata de una base de datos sencilla, por lo que el uso de la no linealidad no era necesario para descubrir relaciones complejas en bases de datos extensas, sino para tratar problemas de datos en el proceso generador de datos. En el capítulo siguiente se examina cómo puede ayudar la modelización no lineal en un caso de base de datos amplia.

Identificación de los indicadores económicos que predicen las recesiones en España

La Gran Recesión provocada por la crisis financiera de 2008 a 2012 tuvo consecuencias económicas que tuvieron un impacto especialmente profundo en España. El ciclo económico español se ha caracterizado por periodos de fuerte crecimiento e intensas recesiones debido a que se trata de una economía con una elevada dependencia del exterior pero al mismo tiempo con un modelo de crecimiento también basado en sectores dependientes del consumo. Sin embargo, la magnitud de la Gran Recesión y sus repercusiones sobre el empleo, el sistema bancario, la vivienda, y la erosión de la confianza económica aumentaron el interés por analizar el ciclo económico español.

En los últimos años se han propuesto diversos indicadores económicos basados de distintas variables económicas, como en Camacho y Doménech (2012) o en Cuevas, Pérez Quirós y Quilis (2017). Al mismo tiempo, se han realizado análisis sobre la sincronización de la actividad económica en España en relación con las economías internacionales, como en Camacho, Caro y López-Buenache (2020), así como análisis de la sincronización a nivel regional como en Gadea-Rivas, Gómez-Loscos y Leiva-León (2019), y análisis de la sincronización de las regiones españolas con el ciclo nacional como en Camacho, Pacce y Ulloa (2018). A pesar de estos esfuerzos, sigue existiendo una notable ausencia de investigaciones centradas en identificar los indicadores más eficaces para anticipar periodos de recesión en la economía española. Esta falta es relevante dado el desfase temporal con el que el Comité de Fechado Cíclico evalúa los picos y valles del ciclo económico.

Recientenmente, se han realizado esfuerzos analíticos para países de importancia económica mundial, como Estados Unidos y Alemania, ambos con capacidad para influir en el ciclo económico mundial con ciertos investigadores utilizado técnicas de árboles de decisión en estos contextos. En concreto, Ng (2014) investigó la capacidad de los árboles de gradient boosting para predecir recesiones económicas en Estados Unidos, mientras que Döpke, Fritsche y Pierdzioch (2017) realizaron un análisis similar para el caso de Alemania. Al mismo tiempo, Ward (2017) exploró la capacidad de los bosques aleatorios (RF por sus siglas en inglés) para

identificar crisis financieras internacionales. Más recientemente, Piger (2020) ha llevado a cabo un análisis comparativo de varios métodos de predicción, incluida la técnica de árboles basados en boosting, para el caso de Estados Unidos y encuentra que ésta última arroja resultados superiores. Así, en este capítulo utilizamos esta técnica para analizar los distintos indicadores para el caso de España utilizando la base de datos MEI de la OCDE, que cuenta con hasta 270 indicadores mensuales, algunos de ellos recogidos desde 1970.

La técnica del gradient boosting en los árboles de decisión permite analizar tanto la capacidad predictiva como la importancia de las variables a la hora de realizar la predicción, e incluso la interacción entre variables a la hora de construir la probabilidad de recesión. El enfoque mostró una capacidad alta para predecir la recesión tanto en horizontes de previsión a tres como a seis meses vista, tanto en el ejercicio dentro de la muestra como en el ejercicio fuera de la muestra. Además, los indicadores adelantados de la tendencia del PIB y de la venta de automóviles junto con los datos de desempleo registrado y de precios parecen ser indicadores clave en la predicción, ayudados por los indicadores de confianza, bursátiles y de tipos de interés. Sin embargo, mientras que los indicadores clave para predecir la Gran Recesión fueron los indicadores financieros y el indicador adelantado de la construcción, para predecir la recesión económica derivada del Covid-19, tanto los precios, como el paro registrado y los indicadores adelantados del PIB y la venta de coches volvieron a ser los indicadores clave.

Por lo tanto, en este capítulo se utiliza una técnica no lineal consolidada para analizar bases de datos complejas e inferir relaciones entre variables. Para profundizar en las ventajas de incorporar cierto grado de no linealidad en el modelado de bases de datos menos complejas, en el capítulo siguiente se demuestra cómo incluso la introducción de sutiles no linealidades en un modelo lineal multivariante puede mejorar la precisión predictiva, superando a menudo el rendimiento de técnicas no lineales más complejas.

Capacidad predictiva de un modelo factorial dinámico para predecir homicidios con arma de fuego

Al igual que ocurre con el ciclo económico, los responsables políticos dependen en gran medida de la información oportuna sobre la evolución de la dinámica de los crímenes para la toma de decisiones que busquen su reducción. De hecho, no sólo los responsables políticos tienen interés en su seguimiento, si no que para agentes económicos como aseguradoras o el sector inmobiliario es también de gran importancia. Como ocurre con los organismos estadísticos al tratar los datos económicos recogidos, los datos sobre delincuencia, debido a la investigación necesaria y a los procedimientos establecidos, están sujetos a un considerable retraso en la publicación de los datos oficiales. Este problema es especialmente relevante en un país muy castigado por la violencia criminal como es Estados Unidos. Como consecuencia de la desigualdad y el acceso a las armas de fuego, el país sufre altos índices de violencia armada que se traducen en consecuencias no sólo sociales sino también económicas. Aunque se observó una reducción de los niveles de delincuencia a principios del siglo XXI, la violencia armada ha aumentado recientemente en Estados Unidos (Gramlich, 2022), lo que se refleja en que la violencia armada se ha convertido en el segundo tema de preocupación para los estadounidenses (Ipsos, 2022).

No obstante, como ocurre con los datos macroeconómicos con los datos de registro o de confianza que aparecen con menor retraso en su publicación, han aparecido también diferentes bases de datos en la descripción de la violencia armada con carácter de publicación más inmediatos: desde bases de datos crowdsource, hasta datos de medios de comunicación o registros de verificación de antecedentes, éstos suponen una cantidad de información previa a la publicación de los datos oficiales que resulta de gran interés para anticiparse a la dinámica a seguir, ya

que el retraso en la publicación por parte del Centro de Control y Prevención de Enfermedades (CDC) de los registros de homicidios con armas de fuego es de hasta 23 meses. Aunque en la literatura se han propuesto modelos autorregresivos univariantes, como los presentados en McDowall (2002), también se han explorado modelos autorregresivos multivariantes, como en Blumstein y Rosenfeld (2008). En esta línea, Cherian y Dawson (2015) introdujeron un modelo ARIMA bivariado que vincula los homicidios con indicadores económicos concurrentes, mientras que Parkin et al. (2020) desarrollaron un modelo VAR que relacionaba la presencia de determinados agentes con las tasas de homicidio. Recientemente, técnicas no lineales como el bosque aleatorio en Berk et al. (2009) o las redes neuronales de memoria corta (LSTM por sus siglas en inglés) en Meskela et al. (2020) y Devi y Kavitha (2021) han mostrado ser también de interés en este contexto, aunque estas técnicas dependen de la riqueza de la base de datos para poder realizar predicciones precisas.

En el contexto de unas fuentes de datos principales cada vez más accesibles, en este tercer capítulo ajustamos un modelo factorial dinámico (Stock y Watson, 1991) a los datos oficiales de los CDC sobre homicidios con armas de fuego, junto con distintas series con datos económicos, de los medios de comunicación o de crowdsourcing, estimados mediante un filtro de Kalman no lineal, con la capacidad de incluir datos de distinta frecuencia y datos ausentes sin actualizar el filtro para dichas observaciones faltantes, siguiendo a Brockwell y Davis (2009). El modelo describe bien la dinámica y también demuestra tener una mayor capacidad predictiva que los modelos autorregresivos paramétricos y los modelos de aprendizaje automático, incluso teniendo en cuenta en éstos también la presencia de esos datos anticipados para sus predicciones, especialmente ante los incrementos registrados como consecuencia de los disturbios tras el caso George Floyd y el repunte durante la pandemia de Covid-19.

Si bien los capítulos hasta este punto profundizan por un lado en la modelización no paramétrica y no lineal, y por otro lado tanto en la aplicación de métodos no lineales a bases de datos complejas como en la inclusión de no linealidades en modelos lineales, la mayor disponibilidad de datos que ha impulsado el uso de técnicas no lineales también se aplica en las técnicas de panel, lo que se analiza a continuación en el último capítulo de esta tesis.

Un modelo de panel dinámico no lineal para explicar las conexiones entre países

Los recientes shocks económicos extraordinarios que suponen la Gran Recesión y la Covid-19 se han producido al mismo tiempo que el aumento del uso de las telecomunicaciones, la globalización del turismo y los niveles de comercio internacional alcanzan niveles sin precedentes. Si bien la globalización ha provocado un cambio de paradigma, su papel en la propagación de las perturbaciones en este mundo más interconectado no está claro, como señala Eickmeier (2007). La transmisión de las perturbaciones es, de hecho, una cuestión aún más relevante para los países de la UEM debido a su particular diseño.

Sin embargo, cuando la sincronización de los ciclos económicos ha sido objeto de investigación, no se ha llegado a un consenso definitivo. Las perspectivas van desde las que afirman la existencia de un único ciclo económico (Giannone, 2010), hasta las que abogan por más de uno (Aguiar-Conrario y Soares, 2009). Ambos puntos de vista se unen parcialmente en Gehringer y Konig (2021), donde apuntan a la idea de que la integración económica impulsó una sincronía que se detuvo con la crisis de deuda. Además, Crespo-Cuaresma y López-Amador (2013) sostienen que de hecho la integración monetaria no aumentó la sincronicidad sino las interdependencias entre un grupo de países. Si bien en este contexto la conectividad se ha inferido recientemente para la UE en Arcabic y Skrinjaric (2021) y para el sector bancario de los países de la UEM en Magkonis y Tsopanakis (2020), la literatura sin embargo carece de

modelos que expliquen tales interdependencias.

Así, en este capítulo, primero derivamos las conectividades (o spillovers) entre la producción industrial mensual de 11 países de la UEM basándonos en la propuesta de Diebold y Yilmaz (2009). A continuación, analizamos la estructura de las conectividades y proponemos un modelo de panel dinámico no lineal con estructura TAR basado en la propuesta de Seo y Shin (2016). El enfoque se inspira en un modelo de gravedad (Anderson y Wincoop, 2003), pero donde el nivel de umbral entre regímenes viene determinado por el índice de conectividad global estimado de la muestra. La introducción tanto de un término dinámico como de un efecto umbral se debe a la naturaleza persistente y asimétrica de la dinámica de las conectividades. De nuestros resultados se desprende que la visión aceptada de la estructura nucleo-periferia no prevalece en la estructura de conectividad obtenida, en línea con los resultados de Matesanz et al. (2017), y también observamos que la conectividad retardada, junto con el turismo y las exportaciones, juegan un papel relevante explicando los spillovers, tanto en el régimen superior como en el inferior definidos por el umbral, mientras que el PIB diferenciado es relevante en el régimen superior y la similitud industrial marginalmente en el régimen inferior. Consecuentemente, la propuesta también proporciona un indicador con dos regímenes con diferentes niveles de conectividad entre países, con implicaciones para los responsables políticos a la hora de establecer medidas para hacer frente a las fluctuaciones económicas, teniendo en cuenta los periodos en los que estas medidas podrían ser más eficaces.

En resumen, esta tesis combina técnicas de modelización no lineales y no paramétricas, análisis de datos complejos y metodologías innovadoras para abordar diversos aspectos de la modelización económica, la previsión y la comprensión de la dinámica económica. Cada capítulo contribuye a la literatura económica en general ofreciendo hallazgos y aplicaciones prácticas tanto para los responsables políticos como para los investigadores y los agentes económicos.

Contents

Chapter 1

Introduction

1.1 — Nonlinear modeling

The primary and enduring method of choice to elucidate relationships, dynamics and, in general, to model the behaviour of both individuals and macroeconomic indicators, has consistently been the linear approach. Its simplicity enables economists to communicate economic ideas and relationships in a clear and concise manner, facilitating their interpretation-an essencial aspect of economic and political decision-making. Additionally, it allows for intuitive analysis, for instance, of elasticities and forecasting.

Nonetheless, seldom reality is driven by linear behavior. Economics is a complex system comprising multiple interdependent variables that may interact nonlinearly, leading to unpredictable or counterintuitive outcomes when traditional linear methods are employed. Numerous instances exist where macroeconomics or microeconomics do not conform to linear behavior. For instance, responses to shocks can be described using nonlinear impulse response functions in case of persistent time series, as seen in Potter (1998) regarding US GNP, or credit can act as a nonlinear propagator economic shocks (Balke, 2000). Furthermore, changes in economic variables can yield non-proportional responses in other variables, as consumption does in response to wealth, income and interest rate shocks (Coskun, Apergis, and Coskun, 2022), or as stock markets response to oil price shocks (Escobari and Sharma, 2020), or GDP does also to them (Karaki, 2017).

Hence, the application of nonlinear techniques in economics has gained significance in recent years, and can provide multiple advantages. The first of all to be highlighted is the higher ability in *accuracy while modeling and forecasting*. Economic dynamics frequently exhibit non straightforward dynamics, such as threshold effects, chaos or second-order derivative effects. Linear techniques, incapable of accommodating these complexities will fall short in terms of accuracy. Therefore, nonlinear models, less prone to suffer in those cases, may lead to higher forecasting accuracy.

While nonlinear modeling can be traced back at least to Hicks (1950) and Goodwin (1951)

as a solution to address the oscilatory time path problem in the lineal approach that for instance appeared in the approach in Samuelson (1939), it was after the development of linear models in time series with the influential paper by Box and Jenkins (1976) that popularized ARIMA models (based on Whittle's seminal work in 1951), and its extension to the multivariate context by Sims (1980), when non-linear models are most developed while *extending* linear approaches. Since the aforemention ARIMA and multivariate models rely on the assumption of linearity and assume symmetry, which seldom are fulfilled, Tong (1978) introduced the concept of threshold effects in the Threshold Autoregressive (TAR) model to take into account asymmetry, later incorporating Anderson and Teräsvirta (1992) a smooth transition in the threshold through the smooth transition autoregressive (STAR) model. For cases where the shift of regime is assumed to be modellable, Tong (1990) proposed the Self-Exciting Threshold Autoregressive (SETAR) models where lagged observations of the dependent variable cause regime shifts, while Hamilton (1989) assumes a Markov process to underlie these shifts. These models have then been extended to a multivariate context, such as the Threshold Vector Autorregresive (TVAR) model (Tsay, 1998; Hubrich and Teräsvirta, 2013), the Vector Smooth Transition Autoregressive Models (Camacho, 2004; Teräsvirta and Yang, 2014), the Multivariate Self-Exciting Threshold Autoregressive (MSETAR) models (Arnold and Günther, 2001), and the Markov Switching Vector Autoregressive (MSVAR) model (Krolzig and Krolzig , 1997).

Nonetheless, the introduction of these more sophisticated dynamics sometimes requires of nonparametric techniques for testing and estimation, less prone to suffer when data present nonlinearities. Boostrap techniques, for instance, have been applied for testing threshold effects in Giannerini, Goracci, and Rahbek (2021), and as estimation technique in the Markov Switching model in Ho (2001), while Genetic Algorithm has been employed as estimation technique in MSETAR models in Baragona and Cucina (2013). Within the same context of extending linear modelling, the EGARCH (Nelson, 1991), the GJR (Glosten, Jagannathan, and Runkle, 1993), or the smooth transition GARCH (Hagerud, 1997; Lanne and Saikkonen, 2005) models are nonlinear extensions of linear GARCH models (Bollerslev, 1986) which failed to correctly capture asymmetric volatility as pointed in Bildirici and Orsen (2014). Additionally, hybrid combinations have been also proposed, as for instance the Fuzzy-EGARCH-ANN model (Mohammed, Aduda and Kubo, 2020) to better capture asymmetries in volatility of financial returns. Last but not least, the widely used Dynamic Factor Models (Stock and Watson, 1991) for business cycle forecasting, as extended in Brockwell and Davis (2009), incorporate nonlinearity by integrating the Kalman Filter without updates when handling missing values, augmenting the initial proposal to a nonlinear framework. Another form of nonlinear extension within these models is the second-order dynamics form proposed in Guerrón-Quintana, Khazanov, and Zhong (2021) and which is estimated through Unscented Kalman Filter.

A second feature to emphasize when delving into nonlinear modeling in time series is its heightened ability to *handle large and intricate data*. In recent times, economic data have increased in size and complexity, resulting in larger datasets with entangled relationships. Certain

nonlinear techniques, such as agent based models and machine learning techiques prove useful in handling such voluminous data. What's more, nonlinear techniques might help to uncover relations that would not be detected through linear techniques. While some of these techniques trace back in time, their utility has become increasingly evident due to the proliferation of data availability, leading to broader adoption in social sciences and economics. For instance, Random Forest (RF) techniques (Breiman, 2001) have recently found application in predicting GDP growth and poverty (Adriansson and Mattsson, 2015; Sohnesen, 2017) employing tools like variable importance measures and partial dependence plots to explore economic relations. Gradient Boosting (Friedman, 2001) has been utilized in predicting failures in the US banking sector (Carmona, Climent and Momparler, 2019) and forecasting real GDP growth (Yoon, 2021). Support Vector Machines (Boser et al. 1992) have been deployed to forecast energy markets prices in Papadimitriou, Gogas, and Stathakis (2014), and Recurrent Neural Networks (Rumelhart et al., 1986) have been applied for forecasting exchange rates in Kuan and Liu (1995), and to forecast nonlinear inflation in Almosova and Andresen (2023).

In fact, when economics, particularly econometrics, emerged as a discipline, the scarcity of both cross-sectional and longitudinal data made linear modelling more appealing than nonlinear modelling. Nonlinear approaches were considered less suitable for extracting information in the face of data scarcity. However, nowadays, the capacity for data collection, the availability of extensive historical data, and enhanced computational capabilities have made nonlinear modeling more alluring and attainable.

Furthermore, and related to the both mentioned higher data availability and the desire to extend linear modeling, in the recent years an increasing literature on *nonlinear panel data models*. The growing storage capacity has made time series data for various individuals readily accessible, making panel data particularly relevant. To incorporate nonlinearities into panel data modeling, Hansen (1999) developsd a panel threshold model, subsequently extended to a smooth transition panel framework by González, Teräsvirta, and van Dijk (2004). Kremer et al. (2013) put forward a threshold dynamic panel model, estimated through GMM, allowing for endogenous regressors but equiring exogeneity of the transition variable. Seo and Shin (2016) generalized this approach allowing endogenous and exogenous regressors but also the transition variable.

In a nutshell, nonlinear techniques in time series analysis offer the potential to model the complexity of economic systems, provide richer descriptions of their dynamics, and enhance predictive accuracy, ultimately leading to more informed decision-making.

1.2 — Contribution

This dissertation endeavors to contribute to this literature through a range of distinct analyses. To provide an overview of these contributions, along with the structure of the dissertation, we can summarize them as follows. In Chapter 2, we propose a non-parametric univariate model to predict ECRI-dated recessions for G7 countries. This is a relevant issue, particularly in the aftermath of the profound economic repercussions of the Covid-19 pandemic, where many forecasting models have experienced diminished accuracy. In Chapter 3, we use a decision tree family model to examine which economic indicators are the best predictors of recessions in the Spanish economy based on a sample of 270 monthly indicators over 50 years. The availability of extensive and comprehensive databases has made it increasingly feasible to utilize techniques capable of identifying intricate and multifaceted relationships. In Chapter 4, we propose the application of a dynamic factor model to gun homicides in the USA, where the model is estimated through a nonlinear Kalman filter due to the presense of missing observations, and compares its performance with other linear and non-linear models. The timeliness of official data publication has significant implications for policymakers, making this research particularly relevant. In Chapter 5, we propose a non-linear dynamic panel model based on a gravity model to explain the cross-country spillovers between euro area countries, previously estimated using the variance decomposition methodology of the prediction errors of a VAR model. Understanding the varying degrees of interconnectedness among countries over time is of paramount importance and carries significant implications for economic and monetary policy.

In addition to their substantive contributions, each of these chapters also offers value in terms of code implementation. Some chapters feature novel model proposals, while others involve the adaptation of models from programming languages such as Gauss or Matlab to R. This transversal dimension adds depth and practicality to the research, making it not only academically meaningful but also accessible and applicable in real-world contexts.

Nonparametric nonlinear univariate modeling to forecast economic recessions

Early warning systems play a key role in the ability of economic agents to make appropriate investment and savings decisions. They also significantly influence the policy decisions of governments and policymakers. Consequently, the academic literature abounds with proposals for early warning systems: from Hamilton's famous article proposing a Markov Switching model (1989), to the STAR approach advocated by Anderson and Terasvirta (1992), and the probit model of Estrella and Mishkin (1998). A little earlier, Wecker (1979) made a proposal for forecasting based on an autoregressive model, which later Camacho and Perez-Quirós (2002) confirmed to have considerable empirical capacity. This is usually done by comparing the periods identified as having a high probability of recession with the periods identified by the National Bureau of Economic Research (NBER) and Economic Cycle Research Institute (ECRI).

Nevertheless, an unprecedented event since economic data have been available, the economic collapse caused by the COVID-19 pandemic and the rapid countercyclical measures implemented by policymakers, caused a fall in economic activity of previously unrecorded magnitude followed by a rebound of similar magnitude, which has caused major adjustment problems in the models mentioned above. Some proposals treat this recession in a unique and special way to adapt parametric approaches (Leiva-León, Perez-Quirós and Rots, 2020; Carriero et al., 2021; Ng, 2021), while the indicators of institutions such as the Econbrowser GDP-based recession indicator index (based on the MS model in Chauvet and Hamilton, 2006) fixes the parameters for the second quarter of 2020 and do not update their calculation in that period, or the St. Louis Fed recession probability indicator (based on the model of Chauvet and Piger, 2008), estimates allowing for a change of model parameters in the aforementioned quarter. Similarly, McGrane (2022) ceased parameter estimation at the end of 2019.

The first chapter of this dissertation contributes to this literature by proposing a non-linear and non-parametric approach to recession forecasting where the proposal does not require modification of the database or the model, which is robust to influential observations. The proposal is constructed on the basis of Wecker's proposal. This extension enables the generation of forecasts for a future time period (h-period ahead) at a given time point t by augmenting the current time series with historical increments found within (h + 1)-dimensional blocks extracted from the time series' past. These forecasted paths are subsequently utilized to calculate the relative frequency of technical recessions as a weighting of the different possible prediction paths, obtained by embedding the series in the symbol space. By means of a Monte Carlo simulation exercise, we analyse the prediction capacity of the proposal in the face of non-linearities in the data against the Wecker approximation and against a Markov Switching model, obtaining a robust performance in our case against the competitors' problems. The empirical exercise is carried out for the G7 countries, obtaining a higher predictive capacity in the in-sample exercise after the incorporated observation derived from the pandemic.

While the first chapter focuses on its contribution to the use of non-linear and non-parametric modeling in univariate time series with data-generating processes featuring non-linearities, it dealt with a simple database so that the use of non-linearity was not needed to uncover complex relationships in extensive databases but to deal with data problems in the data generating process. The subsequent chapter does examines how non-linear modeling can assist in regard to extensive databases.

Identification of the economic indicators that predict recessions in Spain

The Great Recession caused by the financial crisis from 2008 to 2012 had economic consequences that had a particularly profound impact on Spain. Although Spain's economic cycle has been characterized by periods of strong growth and intense recessions due to the fact that it is an economy with a high dependence on external factors, but at the same time with a growth model also based on consumption-dependent sectors. However, the magnitude of the

Great Recession and its repercussions on employment, the banking system, housing, and the erosion of economic confidence heightened interest in analyzing Spain's economic cycle.

In recent years, various economic indicators based on different economic variables have been proposed, such as in Camacho and Doménech (2012) or in Cuevas, Pérez Quirós and Quilis (2017). At the same time, analysis on the synchronization of economic activity in Spain in relation to international economies has been performed, as in Camacho, Caro and López-Buenache (2020), as well as analysis of the synchronization at the regional level as in Gadea-Rivas, Gomez-Loscos and Leiva-León (2019) and analysis of the synchronization of the spanish regions with the national cycle as in Camacho, Pacce and Ulloa (2018). Despite these endeavors, there remains a notable absence of research focused on identifying the indicators that are most effective in anticipating recessionary periods in the Spanish economy. This gap is significant given the time lag with which the Cycle Dating Committee evaluates economic peaks and troughs.

Analytical efforts have been undertaken for countries of global economic importance, such as the United States and Germany, both of which have the capacity to influence the world economic cycle. Researchers have utilized decision tree techniques in these contexts. In particular, Ng (2014) investigated the ability of gradient boosting trees to predict economic recessions in the USA, while Döpke, Fritsche and Pierdzioch (2017) conducted a similar analysis for the case of Germany. At the same time, Ward (2017) explored the ability of random forest to identify international financial crisis. Recently, Piger (2020) conducts a comparative analysis of several predicting methods, including the boosting technique, for the case of the United States and finds that the latter yields superior results. In this chapter, we use this technique to analyze the different indicators for the case of Spain using the OECD MEI database, which has up to 270 monthly indicators, some of them collected since 1970.

The gradient boosting technique makes it possible to analyze both the predictive capacity and the importance of the variables in making the prediction, and even the interaction between variables when constructing the probability of recession. The approach showed very high ability to predict recession in both three and six month ahead forecast horizon, both in sample and in out of sample exercises. Moreover, leading indicators of trend GDP and car sales together with registered unemployment data appeared to be key indicators in the prediction, helped by confidence, stock and interest rates indicators. However, while the key indicators to predict the Great Recession were financial indicators and the leading indicator of construction, to predict economic recession derived from Covid-19 car sales, registered unemployment and leading indicators of GDP were again the key indicators.

Therefore, this chapter makes use of an established nonlinear technique for analyzing complex databases and inferring relationships between variables. To further explore the advantages of incorporating a degree of nonlinearity in modeling less complex databases, the subsequent chapter demonstrates how even the introduction of subtle nonlinearities into a linear model

can enhance predictive accuracy, often surpassing the performance of more complex nonlinear techniques.

Predictive ability of a dynamic factor model predicting homicides with gun

Similar to the economic cycle, policymakers heavily rely on timely insights into evolving crime dynamics. In fact, it is not only political decision-makers who are interested in monitoring it, but also economic agents such as insurers. As is the case with statistical agencies when processing the economic data collected, crime data, due to the necessary research and established procedures, is subject to a considerable delay in official data released. This problem is especially relevant in a country hit particularly hard by criminal violence such as the United States is. As a result of inequality and access to firearms, the country suffers high rates of gun violence that translate into not only social but also economic consequences. While a reduction in crime levels was observed at the beginning of the 21st century, gun violence has recently increased in the United States (Gramlich, 2022), which is reflected in gun violence becoming the second most important issue of concern to Americans (Ipsos, 2022).

Nevertheless, as is also the case with macroeconomic data, with registry or confidence data appearing with little delay in their publication, different databases have appeared in the description of gun violence: from crowdsource databases, to media data or background check registries, these represent a quantity of information ahead of the publication of the data to be published that is of great interest to anticipate dynamics to appear, since the delay in the publication by the Center for Disease Control and Prevention (CDC) of gun homicide records is as long as 23 months. While the literature has proposed univariate autoregressive models, such as those presented in McDowall (2002), it has also explored multivariate autoregressive models, as in Blumstein and Rosenfeld (2008). In this vein, Cherian and Dawson (2015) introduced a bivariate ARIMA model linking homicides to concurrent economic indicators, while Parkin et al. (2020) developed a VAR model relating the presence of certain agents to homicide rates. Recently, non-linear techniques such as random forest in Berk et al. (2009) or short memory neural networks in Meskela et al. (2020) and Devi and Kavitha (2021) have also been shown to be interesting in this context, although these techniques depend on the richness of the database to be able to make accurate predictions.

In the context of increasingly accessible leading data sources, in this third chapter we fit a dynamic factor model (Stock and Watson, 1991) to official CDC data on gun homicides, together with different series with economic, media or crowdsourced data, estimated through a nonlinear Kalman filter, with the ability to include data of different frequency and missing data without updating the filter for missing observations, as in Brockwell and Davis (2009). The model describes the dynamics well and also has a higher capacity than parametric autoregressive and machine learning models, even taking into account the presence of such advance data for its predictions, especially in the face of the increases recorded as a consequence of the riots following the George Floyd case and the spike in the Covid-19 pandemic.

While the initial chapters delve into non-parametric, non-linear modeling and the application of non-linear methods to complex databases and include non-linearities in linear models, the increased availability of data that has driven the use of non-linear techniques also in panel techniques, discussed in the last chapter of this thesis below.

A dynamic nonlinear panel model to explain cross-country connectedness

The recent extraordinary economic shocks, spanning the Great Recession and the Covid-19 recession, have occurred concurrently with the increased use of telecommunications, the globalization of tourism, and record-high levels of international trade. While globalisation has brought about a paradigm shift, its role in the propagation of shocks within this more interconnected world is not clear, as Eickmeier (2007) points out. The transmission of shocks is in fact an issue of even greater relevance for EMU countries because of their particular design.

Although the synchronization of economic cycles has been a subject of thorough investigation, yet it has failed to yield a definitive consensus. Perspectives range from those asserting the existence of a single business cycle (Giannone, 2010), to those views advocating for more than one (Aguiar-Conrario and Soares, 2009). Both views are partially joined in Gehringer and Konig (2021) where they point to the idea that economic integration drove a synchronicity that stopped with the debt crisis. Furthermore, Crespo-Cuaresma and Lopez-Amador (2013) argue that monetary integration did not increase synchronicity but interdependencies among a group of countries. While connectivity has recently been inferred for EU in Arčabić and Škrinjarić (2021) and for EMU countries' banking sector in Magkonis and Tsopanakis (2020), however, there is a lack of modeling to explain such interdependencies in the literature.

Hence, in this chapter, we first derive the cross-country connectivities based on the Diebold and Yilmaz (2009) proposal for the monthly industrial production of 11 countries from the EMU. Then, we analyze the structure of the connectivities, and propose a nonlinear dynamic panel model with TAR structure based on the Seo and Shin (2016) proposal. The approach is inspired by a gravity model (Anderson and Wincoop, 2003), where the treshold level between regimes was determined by the estimated global connectivity index of the sample. The introduction of both a dynamic term and a threshold effect is due to the persistent and asymmetric nature of the spillovers' dynamics. We observed from our findings that the accepted core-periphery view was not prevailing in the connectivity structure obtained, in line with the results in Matesanz et al. (2017), and we also obtained that the lagged spillover, together with tourism and exports play a relevant role explaining the spillovers in both the upper and lower regimes defined by the treshold, while the differentiated GDP was relevant in the upper and the industrial similarity marginally in the lower regime. The proposal also provides therefore an indicator with two regimes with different levels of connectivity between countries, with implications for policymakers when establishing measures to address economic fluctuations, taking into account the periods where these measures could be most effective.

In summary, this dissertation combines non-linear and non-parametric modeling techniques, complex data analysis, and innovative methodologies to address various aspects of economic modeling, forecasting, and understanding economic dynamics. Each chapter contributes to the broader economic literature by offering unique insights and practical applications for policymakers and researchers alike.

1.3 — References

Adriansson, N., and Mattsson, I., 2015. Forecasting GDP Growth, or How Can Random Forests Improve Predictions in Economics? Bachelor Thesis. Uppsala, Sweden: Uppsala University.

Aguiar-Conraria, L., and Soares, M. J., 2009. Business cycle synchronization across the Euro area: A wavelet analysis. NIPE, Working Paper No. 8/2009. Universidado De Minho.

Almosova, A., and Andresen, N., 2023. Nonlinear inflation forecasting with recurrent neural networks. *Journal of Forecasting*, 42(2), 240-259.

Anderson H., and Teräsvirta T., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregression models. *Journal of Applied Econometrics* 7: S199-136.

Anderson, J. E., and van Wincoop, E., 2003. Gravity with Gravitas: A Solution to the Border Puzzle *American Economic Review*, 93, 170-192.

Arčabić, V., and Škrinjarić, T., 2021. Sharing is caring: Spillovers and synchronization of business cycles in the European Union. *Economic Modelling*, 96, 25-39.

Arnold, M., and Günther, R., 2001. Adaptive parameter estimation in multivariate self-exciting threshold autoregressive models. *Communications in Statistics-Simulation and Computation*, 30(2), 257-275.

Balke, N. S., 2000. Credit and economic activity: credit regimes and nonlinear propagation of shocks. *Review of Economics and Statistics*, 82(2), 344-349.

Baragona, R., and Cucina, D., 2013. Multivariate self-exciting threshold autoregressive modeling by genetic algorithms. *Jahrbücher für Nationalökonomie und Statistik*, 233(1), 3-21.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., and Ahlman, L., 2009. Forecasting murder within a population of probationers and parolees: a high stakes application of statistical learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172 (1), 191–211.

Bildirici, M., and Ersin, Ö. Ö., 2014. Nonlinearity, volatility and fractional integration in daily oil prices: Smooth transition autoregressive ST–FI (AP) GARCH models. *Romanian Journal of Economic Forecasting*, 3, 108-135.

Blumstein, A., and Rosenfeld, R., 2008. Factors contributing to US crime trends. In *Understanding crime trends: Workshop report* (Vol. 2, pp. 13–44). National Academies Press.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.

Box, G.E.P. and Jenkins, G.M., 1976. *Time Series Analysis Forecasting and Control*. 2nd Edition, Holden-Day, S. Francisco.

Boser, B. E., Guyon, I. M., and Vapnik, V. N., 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).

Breiman, L., 2001. Random forests. *Machine learning*, 45, 5-32.

Brockwell, P. J., and Davis, R. A., 2009. *Time series: theory and methods. Springer Science & Business Media*.

Camacho, M., 2004. Vector smooth transition regression models for US GDP and the composite index of leading indicators. *Journal of Forecasting*, 23(3), 173-196.

Camacho, M., Caro, A., and López-Buenache, 2020. The two-speed Europe in business cycle synchronization. *Empirical Economics*, 59(3), 1069-1084.

Camacho, M., and Doménech, R., 2012. MICA-BBVA a factor model of economic and financial indicators for short-term GDP forecasting. SERIEs: *Journal of the Spanish Economic Association* 3: 475-497.

Camacho, M., Pacce, M., and Ulloa, C., 2018. Regional business cycle phases in Spain. *Estudios de Economia Aplicada* 36: 875-896.

Camacho, M., and Perez Quirós, G. 2002. This is what the leading indicator lead. *Journal of Applied Econometrics* 17: 61-80.

Carmona, P., Climent, F., and Momparler, A., 2019. Predicting failure in the US banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, 61, 304-323.

Carriero, A., Clark, T., Marcellino, M., and Mertens, E. 2021. Measuring uncertainty and its effects in the COVID-19 era. CEPR Discussion Papers 15965.

Chauvet, M., and Hamilton, J. 2006. Dating Business Cycle Turning Points. In Milas, C., Rothman, P., van Dijk, D., and Wildasin, E. (eds) *Nonlinear Time Series Analysis of Business Cycles*. Emerald Group Publishing Limited: Bingley.

Chauvet, M., and Piger J. 2008. A comparison of the real-time performance of business cycle dating methods. *Journal of Business and Economic Statistics* 26: 42-49.

Cherian, J., and Dawson, M., 2015. RoboCop: Crime classification and prediction in San Francisco. *Forest*, 15 .

Coskun, Y., Apergis, N., and Coskun, E. A., 2022. Nonlinear responses of consumption to wealth, income, and interest rate shocks. *Empirical Economics*, 63(3), 1293-1335.

Crespo-Cuaresma, J., and Fernández-Amador, O., 2013. Business cycle convergence in EMU: A first look at the second moment. *Journal of Macroeconomics*, 37, 265-284.

Cuevas, A., Pérez Quirós, G., and Quilis, E., 2017. Integrated model of short-term forecasting of the Spanish Economy (Mipred Model). *Revista de Economía Aplicada* 25: 5-25.

Devi, J. V., and Kavitha, K., 2021. Automating time series forecasting on crime data using RNNLSTM. *International Journal of Advanced Computer Science and Applications*, 12 (10).

Diebold, F., and Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. *Economic Journal*, 119, 158-171.

Döpke, J., Fritsche, U., and Pierdzioch, C., 2017. Predicting recessions with boosted regression trees. *International Journal of Forecasting* 33: 745-759.

Eickmeier, S. 2007. Business Cycle Transmission from the US to Germany-A Structural Factor Approach. *European Economic Review*, 51: 521–551.

Escobari, D., and Sharma, S., 2020. Explaining the nonlinear response of stock markets to oil price shocks. *Energy*, 213, 118778.

Estrella, A. and Mishkin, F., 1998. Predicting US recessions: financial variables as leading indicator. *Review of Economic and Statistics* 80: 45-61.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.

Gadea-Rivas, M., Gomez-Loscos, A., and Leiva-León, D., 2019. Increasing linkages among European regions. The role of sectoral composition. *Economic Modelling* 80: 222-243.

Gehringer, A., and König, J., 2021. Recent patterns of economic alignment in the european (monetary) union. *Journal of Risk and Financial Management*, 14(8), 362.

Giannerini, S., Goracci, G., and Rahbek, A., 2023. The validity of bootstrap testing for threshold autoregression. *Journal of Econometrics*

Giannone, D., Lenza, M. and Reichlin, L., 2010. Business Cycles in the Euro Area. *Europe and the Euro*. Ed. Alesina, A. and Giavazzi, F., 141–167. The University of Chicago Press.

Glosten, L. R., Jagannathan, R., and Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5), 1779-1801.

González, A., Teräsvirta, T., and van Dijk, D., 2004. Panel smooth transition regression model and an application to investment under credit constraints. Unpublished manuscript, Stockholm School of Economics.

Goodwin, R. M., 1951. The nonlinear accelerator and the persistence of business cycles. *Econometrica: Journal of the Econometric Society*, 1-17.

Gramlich, J., 2022. What the data says about gun deaths in the US. *Pew Research Center*.

Guerrón-Quintana, P. A., Khazanov, A., and Zhong, M., 2021. Nonlinear Dynamic Factor Models. Unpublished manuscript.

Hagerud, G. E., 1997. *A smooth transition ARCH model for asset returns*. Stockholm School of Economics, the Economic Research Inst.

Hamilton, J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle *Econometrica* 57: 357-384.

Hansen, B. E., 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics*, 93(2), 345-368.

Hicks, J., 1950. *A Contribution to the Theory of the Trade Cycle*. Clarendon Press, Oxford.

Ho, T. W., 2001. Finite-sample properties of the bootstrap estimator in a Markov-switching model. *Journal of Applied Statistics*, 28(7), 835-842.

Hubrich, K., and Teräsvirta, T., 2013. Thresholds and Smooth Transitions in Vector Autoregressive Models. In *VAR models in macroeconomics–New developments and applications: Essays in honor of Christopher A. Sims* (pp. 273-326). Emerald Group Publishing Limited.

Ipsos., 2022. *FiveThirtyEight/Ipsos 2022 Election Tracking Survey*. (https://www.ipsos.com/en-us/ news-polls/FiveThirtyEight-2022-midterm-election)

Kremer, S., Bick, A., and Nautz, D., 2013. Inflation and growth: new evidence from a dynamic panel threshold analysis. *Empirical Economics*, 44, 861-878.

Krolzig, H. M., and Krolzig, H. M., 1997. The markov-switching vector autoregressive model. *Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis*, 6-28.

Kuan, C. M., and Liu, T. (1995). Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*, 10(4), 347-364.

Lanne, M., and Saikkonen, P., 2005. Non-linear GARCH models for highly persistent volatility. *The Econometrics Journal*, 8(2), 251-276.

Leiva-Leon, D., Perez-Quiros, G., and Rots, E. 2020. Real-time weakness of the global economy. European Central Bank Working Paper No 2381.

Karaki, M. B., 2017. Nonlinearities in the response of real GDP to oil price shocks. *Economics Letters*, 161, 146-148.

Magkonis, G. and Tsopanakis, A., 2020. The financial connectedness between eurozone core and periphery: a disaggregated view. *Macroeconomic Dynamics*, 24(7), 1674-1699.

Matesanz Gomez, D., Ferrari, H. J., Torgler, B., and Ortega, G. J., 2017. Synchronization and diversity in business cycles: a network analysis of the European Union. *Applied Economics*, 49(10), 972-986

McDowall, D., 2002. Tests of nonlinear dynamics in US homicide time series, and their implications. *Criminology*, 40 (3), 711–736.

McGrane, M., 2022. A Markov-switching model of the unemployment rate. Congressional Budget Office No. 2022-05.

Meskela, T. E., Afework, Y. K., Ayele, N. A., Teferi, M. W., and Mengist, T. B., 2020. Designing time series crime prediction model using long short-term memory recurrent neural network. *International Journal of Recent Technology and Engineering*, 9(4), 402-405.

Mohammed G.T., Aduda, J.A., and Kube, A.O., Improving Forecasts of the EGARCH Model Using Artificial Neural Network and Fuzzy Inference System. *Hindawi Journal of Mathematics*; c2020. p. 1-14.

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.

Ng, S., 2014. Boosting recessions. *Canadian Journal of Economics* 47: 1-34.

Ng, S., 2021. Modeling macroeconomic variations after COVID-19. NBER Working Papers No. 29060.

Papadimitriou, T., Gogas, P., and Stathakis, E., 2014. Forecasting energy markets using support vector machines. *Energy Economics*, 44, 135-142.

Parkin, W. S., Bejan, V., and Hickman, M. J., 2020. Police, public and community violence: Exploring the relationships between use of deadly force, law enforcement killed, and homicide rates in the United States. *Criminology, Criminal Justice Law and Society*, 21 , 1.

Piger, J., 2020. Turning points and classification. In *Macroeconomic forecasting in the era of big data: Theory and application*, Peter Fuleky (editor). Springer International Publishing.

Potter, S. M., 2000. Nonlinear impulse response functions. *Journal of Economic Dynamics and Control*, 24(10), 1425-1446.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1986. Learning representations by back-propagating errors. nature, 323(6088), 533-536.

Samuelson, P., 1939. Interaction between the Multiplier Analysis and the Principle of Acceleration. *Review of Economic Statistics*, 4, 75-78.

Seo, M. H., and Shin, Y., 2016. Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics*, 195(2), 169-186.

Sims, C. A., 1980. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1-48.

Sohnesen, T. P., and Stender, N., 2017. Is random forest a superior methodology for predicting poverty? An empirical assessment. *Poverty and Public Policy*, 9(1), 118-133.

Stock, J. H., and Watson, M. W., 1991. A probability model of the coincident economic indicators. In *Leading economic indicators: New approaches and forecasting records*. Cambridge University Press.

Teräsvirta, T., and Yang, Y., 2014. *Specification, estimation and evaluation of vector smooth transition autoregressive models with applications* (No. 2014062). Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Tong, H., 1978. *On a Threshold Model in Pattern Recognition and Signal Processing*, Edited by: Chen, C. H. Amsterdam: Sijhoff and Noordhoff.

Tong, H., 1990. *Non-linear time series: a dynamical system approach*. Oxford university press.

Tsay RS., 1998. Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93: 1188–1202.

Ward, 2017. Spotting the danger zone: Forecasting financial crisis with classification tree ensembles and many predictors. *Journal of Applied Econometrics* 32: 359-378.

Wecker, W., 1979. Predicting the turning points of a time series. *Journal of Business* 52: 35-50.

Whittle, P., 1951. *Hypothesis testing in time series analysis*. Thesis, Uppsala University, Almqvist and Wiksell, Uppsala.

Yoon, J., 2021. Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. *Computational Economics*, 57(1), 247-265.

Chapter 2

Nonparametric nonlinear univariate modeling to forecast economic recessions

2.1 — Introduction

Early detection of changes in business cycle phases is crucial for consumption, investment, savings, and production decisions made by economic agents, as well as for making monetary and fiscal policies. Since phase changes are officially recognized long after they start, developing early warning mechanisms to forecast recessions has been a long-standing quest for academics, market practitioners, and policymakers.

To provide timely early warning of an approaching recession, academics usually rely on nonlinear parametric methods that produce probabilistic statements of future phase changes from business cycle indicators, such as the growth rates of quarterly real Gross Domestic Product (GDP). To name only a few, Estrella and Mishkin (1998) used a probit model, Hamilton (1989) developed a Markov-switching autoregressive (MSAR) specification, and Teräsvirta and Anderson (1993) proposed a Smooth Transition Autoregressive (STAR) model.

In addition, Wecker (1979) offered a heuristic solution to compute inferences of future recessions from linear autoregressive (AR) models based on Monte-Carlo simulations of forecast paths. Considering a technical recession (two consecutive quarters of decline in the GDP) as a recession event, the method consists of computing the relative frequency of technical recessions across the simulated forecasts. Despite its simplicity, Hamilton and Perez-Quirós (1996) and Camacho and Perez-Quirós (2002) showed the considerable empirical reliability of this approach to provide forecasts of US recession probabilities.

One way to assess the goodness of fit of recession forecast models is to examine their ability to identify official recessions as determined by national Dating Committees, such as the committee of the National Bureau of Economic Research (NBER) for the US economy. Regard-

less of how successful the parametric methods have been in the past, in 2020, the parametric dating methods were exposed to unprecedented atypical data that challenged their ability to provide reliable probability forecasts of future recessions. In 2020, due to the economic collapse caused by the COVID-19 pandemic and the rapid countercyclical measures implemented by policymakers, most industrialized countries recorded the sharpest fall and the largest rebound in quarterly GDP since records began. In this chapter, we show that these leverage points have dramatically altered the state of affairs in performing business cycle inferences.[1]

To overcome this drawback, the empirical approaches used to compute forecasts of recession probabilities with parametric models rely on the shortcut of manipulating the sample to estimate the model parameters. One example is the Econbrowser GDP-based recession indicator index, which estimates recession probabilities by applying a methodology based on the Markov-switching model developed by Chauvet and Hamilton (2006). To keep the index working after the dramatic drop in the second quarter of 2020, the parameters of the Markov-switching model were not estimated but fixed with the values of the estimated parameters using data only up to the first quarter of 2020. In the same vein, McGrane (2022) obtained the post-COVID recession probabilities from a Markov-switching model estimated with data only up to 2019. On a monthly basis, the recession probability index maintained by the St. Louis Fed and documented by Chauvet and Piger (2008), is computed from a Markov-switching model that allows a change in model parameters associated with the period from March 2020 to July 2020.

The aim of this chapter is to introduce a new approach to compute h-step ahead predictive probabilities of future recessions that does not require manipulating the sample of the database because the method is robust to influential points and other data irregularities, such as structural breaks, heteroskedasticity and Autoregressive Conditional Heteroskedasticity (ARCH). Specifically, we propose a nonparametric extension of Wecker's method that generates h-period ahead forecast paths at time t by adding to the time series under consideration at time t the past increments occurred in the set of $(h+1)$-dimensional blocks that can be extracted from the past of the time series. Then, the forecast paths can be used to compute the relative frequency of technical recessions.

As we will show below, this strategy implies assuming equal weigths for all the past increments, which seems unreliable in practice. For example, past blocks of the time series referring to recovery periods exhibit upward trends that would hardly ever occur when time t refers to a downturn. To overcome this drawback, we embed all the forecast paths into a symbolic space and derive the expressions required to compute the probability of occurrence of each of the symbols at the time of the forecast. Thus, we compute the relative frequency of a technical recession across the forecast paths by weighting each path differently according to

[1] Although not in the context of forecasting recessions, Lenza and Primiceri, 2020, Leiva-Leon, Perez-Quiros and Rots, 2020, Antolin-Diaz, Drechsel and Petrella, 2021, Carriero at al., 2021, and Ng, 2021, have recently investigated some ways to handle the unique features of the COVID-19 recession in time-series forecasting with parametric approaches.

the probability of the corresponding symbol occurring at time t.

The advantage of forecasting recession probabilities in this way compared to linear and nonlinear parametric approaches is twofold. Firstly, the method is nonparametric so there is no need to make any assumption in a specific dynamic model for the given population. In addition, performing of recession probabilities to provide accurate business inferences does not depend on the estimates of the model parameters, which tend to be unstable under structural breaks and large outliers. Secondly, the impact of extreme values appearing in some past blocks of the time series, like those observed in the pandemic period, are expected to average out when computing the weighted relative frequency of the technical recessions and their impact on forecasting performance becomes negligible.

By conducting several Monte Carlo experiments designed to capture standard data problems that characterize economic data sets, we evaluate the finite-sample performance of the proposed algorithm to predict recession probabilities. To assess its forecasting performance, we analyze Receiver Operating Characteristic (ROC) curves, Brier Scores, and Cohen's Kappa coefficients. In absence of data problems, we use these statistics in an out-of-sample forecasting scenario to show that the nonparametric approach developed in this chapter behaves similarly to Markov-switching and Wecker's approaches in one-period forecasting but its relative improvement over parametric approaches consistently increases with the forecasting horizon. In the presence of influential observations and structural breaks, and when the errors are heteroskedastic or present ARCH dynamics, the nonparametric approach outperforms the parametric models.

Finally, we evaluate the ability of our nonparametric method to compute accurate in-sample forecasts of recession probabilities of future recessions (captured by NBER and ECRI recession dates) from national GDP growth rates in the G7 countries. Using pre-pandemic data, the Markov switching approach slightly outperforms the other approaches at one-period forecasting, although there are no sizeable differences with the nonparametric proposal as the forecasting horizon increases. Undoubtedly, the best-performing model is the nonparametric approach when the extreme values of GDP growth rates observed in 2020 are included in the sample because its forecasts of recession probabilities are barely affected by these influential observations.

The chapter is organized as follows. Section 2 introduces the nonparametric approach to compute forecasts of recession probabilities. Section 3 shows the results of the Monte Carlo simulations. Section 4 applies the models to forecast recession probabilities in the G7 countries from quarterly GDP growth rate data. This section highlights the estimation problems faced by parametric approaches when the extreme figures observed in 2020 are included in the data set. Section 5 concludes and outlines some further research lines. This is followed by the references, appendices, tables and figures referred to in the chapter.

2.2 — Robust probabilistic recession statements

Based on a novel extension of Wecker's (1979) proposal, this section describes a new procedure to compute h-period forecasts of recession probabilities. For clarity, we first describe the linear approach and then present our nonparametric extension.

2.2.1. The linear approach

The approach proposed by Wecker (1979) offers a heuristic solution to compute business cycle inferences with linear autoregressive models based on Monte-Carlo simulations. Let $\{y_1, \ldots, y_t\}$ be the observed values of a stationary and ergodic time series, y_t, and let $\{\hat{y}_{t+1}, \ldots, \hat{y}_{t+h}\}$ be the predictions of its uncertain future values $\{y_{t+1}, \ldots, y_{t+h}\}$.

To adapt the method to our context, let y_t be the growth rates of the seasonally adjusted real GDP series for a given country. We rely on the popular definition of a technical recession to state the occurrence of a recession, which requires the GDP to fall for at least two consecutive quarters.[2] To obtain a probabilistic statement about the event of a recession, we define z_t as the sequence of indicator variables that indicate a recession at t, whose outcomes rely on the time series y_t according to the rule

$$z_t(y_{t-1}, y_t) = \begin{cases} 1 & \text{if } y_{t-1} < 0 \text{ and } y_t < 0 \\ 0 & \text{otherwise} \end{cases}. \tag{2.1}$$

Thus, the rule identifies a recession after two (or more) successive declines.

To estimate the h-step-ahead forecast of the probability of a recession, the vector of present and future values of the time series $Y_h(t+1) = \{y_t, y_{t+1}, \ldots, y_{t+h-1}, y_{t+h}\}$, for $h \geq 1$, must be estimated. Assuming that the data-generating process is a univariate autoregressive Gaussian model, sample paths of the future values of the time series can be repeatedly generated. Concretely, one can draw a number M of vectors of forecasts $\{(\hat{y}^m_{t+1}, \ldots, \hat{y}^m_{t+h})\}_{m=1}^M$ from $N(\mu_t, Q)$, where explicit forms of the mean and covariance matrix for different forecasting horizons are derived in Appendix A. This leads to $(h+1)$-dimensional forecast paths

$$\hat{Y}^m_{h+1}(t) = (y_t, \hat{y}^m_{t+1}, \ldots, \hat{y}^m_{t+h}), \tag{2.2}$$

where $m = 1, \ldots, M$ and we set $\hat{y}^m_t = y_t$.

Then, using the rule of two consecutive periods of declining stated in (2.1), we can compute M realizations of the indicator variable $\mathcal{Z}^M_{t+h} = \{z_{t+h}(\hat{y}^m_{t+h-1}, \hat{y}^m_{t+h})\}_{m=1}^M$. Notice that, since the dis-

[2] Among others, this definition has been used by Hamilton and Perez-Quiros (1996) and by Camacho and Perez-Quiros (2002) to compute business cycle inferences from linear autoregressive models.

tribution of future values of the time series conditioned to its past values, $g_y(y_{t+1},\dots,y_{t+h}|y_1,\dots,y_t)$, can be approximated by the empirical distribution of the generated forecast paths $\hat{Y}^m_{h+1}(t)$, the distribution of z_{t+h} can also be approximated by the empirical distribution of $\mathcal{Z}^M_{t+h}$, as stated in Wecker (1979). Therefore, the sample mean of this empirical distribution is taken to be a forecast of the probability that the economy will be in recession at date $t+h$

$$P_L(z_{t+h}=1)=\frac{1}{M}\sum_{m=1}^{M} z_{t+h}(\hat{y}^m_{t+h-1},\hat{y}^m_{t+h}), \tag{2.3}$$

for any look-ahead horizon h.

From this expression, it becomes clear that the ability of P_L to detect future recessions depends crucially on the performance of the forecasting model used to compute reliable forecasts of $(y_{t+1},\dots,y_{t+h})$. In this context, it is evident that model misspecification, structural breaks, or extreme values inducing instability in the autoregressive parameters will negatively impact the performance of P_L.

2.2.2. Nonparametric forecasts

As in the linear approach, our proposal to infer whether an economy will be in recession with forecast horizon h, requires estimating $Y_{h+1}(t)=\{y_t,y_{t+1},\dots,y_{t+h-1},y_{t+h}\}$. However, instead of generating forecast paths from parametric autoregressive models, we rely on simulating nonparametric forecasts by embedding the time series $\{y_t\}_{t=1}^T$ in a $(h+1)$-dimensional space by computing the histories

$$Y_{h+1}(\tau)=(y_\tau,y_{\tau+1},\dots,y_{\tau+h-1},y_{\tau+h}), \tag{2.4}$$

where $\tau=1,\dots,T-h$. Each of these vectors summarizes the behavior of the time series in the neighborhood of τ, accounting for the value of the stationary time series at τ and the subsequent steps $\tau+1,\dots,\tau+h$.

For each τ, we use $Y_{h+1}(\tau)$ to generate realizations of the forecast of $Y_{h+1}(t)$ as follows:

$$\hat{Y}^\tau_{h+1}(t)=(y_t,\hat{y}^\tau_{t+1},\dots,\hat{y}^\tau_{t+h}), \tag{2.5}$$

where $\hat{y}^\tau_{t+k}=y_t+(y_{\tau+k}-y_\tau)$ is the τ-th generation of the forecast y_{t+k}, for $k=1,2,\dots,h$ and $\tau=1,2,\dots,t-h$. In this proposal, the τ-th forecast of y_{t+k}, given by $\hat{y}^\tau_{t+k}$, is the value of y_t plus the increment produced in the time series in the following k step starting at a given period τ.

Apart from it being stationary and ergodic, we do not require further assumptions about the data-generating process of y_t, its population probability distribution. In this case, it is straightforward to show that $E(\hat{y}^\tau_{t+k})=E(y_t)$. In addition, the increments in the time series, $y_{t+k}-y_t$ as well as $y_{\tau+k}-y_\tau$, are stationary because the time series is also stationary, which implies that they are realizations of the same distribution distribution that depends on k but

not on t. This supports our approach as a natural way to perform the different forecasts.

Using the $t-h$ forecast paths $\hat{Y}^{\tau}_{h+1}(t)$, we can generate a sequence of indicators of a technical recession $z_{t+h}(\hat{y}^{\tau}_{t+h-1}, \hat{y}^{\tau}_{t+h})$ for $\tau = 1, \ldots, t-h$, whose empirical distribution approximates the distribution of z_{t+h}. Thus, as a natural extension of the linear approach, the probability of a recession at $t+h$ can be estimated by

$$P(z_{t+h} = 1) = \frac{1}{t-h} \sum_{\tau=1}^{t-h} z_{t+h}(\hat{y}^{\tau}_{t+h-1}, \hat{y}^{\tau}_{t+h}), \tag{2.6}$$

which is the frequency of a technical recession across the nonparametric simulations of the forecast path.

It is worth emphasizing that this extension of the linear approach to producing probabilistic statements for the occurrence of recessions could lead to invalid inferences. Notice that in the linear autoregressive case, the forecasts of the variable of interest $y_{t+1}, ..., y_{t+h}$, account for the inertia of the time series since they are computed using linear combinations of past values of the variable, with the recent past having the highest weights.

By contrast, our nonparametric projections do not consider the typical business-cycle inertia of GDP growth rates when developing the short-term forecasts at t. In fact, expression (2.6) computes the probability of a recession at $t+h$ by averaging the $t-h$ estimates of the indicator variable z, which assigns equal weights to all the indicators of recession $z_{t+h}(\hat{y}^{\tau}_{t+h-1}, \hat{y}^{\tau}_{t+h})$, regardless of the neighborhoods of the time series at t and τ. Thus, when the economy is in the middle of an expansion at t, the method would assign the same weight to $\hat{Y}^{\tau}_{h+1}(t)$ regardless of whether τ refers to an expansionary or to a recessionary period.

The following section proposes a modification of (2.6) that overcomes this drawback using a weighting algorithm based on symbolic dynamics.

2.2.3. Symbolic dynamics based weights

Symbolic dynamics involves the simple process of labeling each of the $(h+1)$-history $Y_{h+1}(t) = (y_t, y_{t+1}, \ldots, y_{t+h})$, for $t = 1, \ldots, T-h$, with a symbol. Thus, instead of following the trajectory of the time series point by point, one only keeps recording the alternation of the symbols. According to Collet and Eckmann (2009), the evolution of the symbols can capture the complete description of the dynamic system[3].

The proposed symbolization is as follows. Let S_{h+1} be the symmetric group of order $(h+1)!$, that is the group formed by all the permutations of length $h+1$ of the elements in the set $\{0, 1, 2, \ldots, h\}$. An element of this group, namely $\pi = (i_0, i_1, ..., i_h) \in S_{h+1}$, is called a

[3]Some other applications of symbolic dynamics in economics are Tino et al. (2000), Matilla, Ruiz, and Dore (2014), Hou et al. (2017) and Camacho, Romeu, and Ruiz (2021).

symbol. The symbolization procedure consists on assigning a unique permutation that sorts out its entries from the smallest to the largest to any $(h+1)$-tuple $Y_{h+1}(t)$. Formally, this procedure maps $Y_{h+1}(t)$ to the unique permutation $\pi = (i_0, i_1, ..., i_h) \in S_{h+1}$, satisfying the following two conditions:

$$y_{t+i_0} \leq y_{t+i_1} \leq ... \leq y_{t+i_h}, \tag{2.7}$$

$$i_{s-1} < i_s \; if \; y_{t+i_{s-1}} = y_{t+i_s}. \tag{2.8}$$

The first condition imposes an ordinal pattern and the second is a technical condition that guarantees the uniqueness of the symbol in the case of equal values, which theoretically has zero probability of occurring in the case of continuous distributions. Thus, symbolic dynamics converts the sequences of $(h+1)$-histories $\{Y_{h+1}(t)\}_{t=1}^{T-h}$ into sequences of ordinal patterns labeled with symbols $\{\pi(t)\}_{t=1}^{T-h}$.

As a quick example, consider the time series $\{y_t\} = \{5, 3, 2, 1, 8, 9, 3, 4, 5, 2\}$ of length $T = 10$. For a forecasting horizon $h = 2$, we can obtain eight 3-histories

$$\{Y_3(t)\}_{t=1}^{8} = \{(5,3,2), (3,2,1), (2,1,8), (1,8,9), (8,9,3), (9,3,4), (3,4,5), (4,5,2)\}. \tag{2.9}$$

Using the integers $\{0, 1, 2\}$ to construct the symbols, the set of potential symbols is

$$S_3 = \{(0,1,2), (0,2,1), (1,0,2), (1,2,0), (2,0,1), (2,1,0)\}. \tag{2.10}$$

Now, symbolic dynamics yields the symbolized series as

$$\{\pi(t)\}_{t=1}^{8} = \{(2,1,0), (2,1,0), (1,0,2), (0,1,2), (1,2,0), (2,0,1), (0,1,2), (1,2,0)\}. \tag{2.11}$$

It is worth emphasizing that symbol (0,1,2) will be associated with increasing patterns (expansions) in the time series, while symbol (2,1,0) will typically refer to decreasing dynamics (recessions).

In addition, the probability at t of symbol $\pi \in S_{h+1}$, which we call P_{π}^{t}, can be computed analytically. Based on an extension of Abd Alla (2004), Appendix B shows the expressions of these forecasts for $h = 1, 2, 3$. These forecasts of symbol probabilities can be used to improve the accuracy of the recession probability forecasts stated in (2.6).

To this end, let τ be any time period smaller or equal to $t - h$, and denote with $\pi(\tau)$ the symbol associated with $(h+1)$-history $Y_{h+1}(\tau)$. Let $P_{\pi(\tau)}^{t}$ be the probability that symbol $\pi(\tau)$

appears at t.[4] We propose the nonparametric forecast at t of a recession occurring at $t+h$ as

$$P_{NP}(z_{t+h}=1)=\frac{\sum_{\tau=1}^{t-h} z_{t+h}(\hat{y}^{\tau}_{t+h-1},\hat{y}^{\tau}_{t+h})P^{t}_{\pi(\tau)}}{\sum_{\tau=1}^{t-h} P^{t}_{\pi(\tau)}}. \tag{2.12}$$

This expression implies that the contribution to $P_{NP}(z_{t+h}=1)$ of the τ-th recession indicator is weighted by the probability that the ordinal pattern of $(h+1)$-history $Y_{h+1}(\tau)=(y_{\tau},y_{\tau+1},\ldots,y_{\tau+h})$ (i. e. symbol $\pi(\tau)$) would occur at t, where the weights are normalized to add up to one.[5]

It is easy to check that forecasting probabilities of recession with the weighted average of the recession indicators as in (2.12) overcomes the drawback of (2.6). Let us denote with π_E those symbols associated with expansions and with π_R those associated with recessions. If we assume that the economy is in the middle of an expansion at t, then probability $P^{t}_{\pi_E}$ should be much higher than probability $P^{t}_{\pi_R}$. Therefore, if period τ is in the middle of a recession, the corresponding $(h+1)$-history, $Y_{h+1}(\tau)$, will be associated with $\pi(\tau)=\pi_R$ and the weight of the recessionary indicator $z_{t+h}(\hat{y}^{\tau}_{t+h-1},\hat{y}^{\tau}_{t+h})$ appearing in the numerator of (2.12) will be as low as $P^{t}_{\pi_R}$.

2.3 — Monte Carlo simulation

In this section, we set up several Monte Carlo experiments to assess the finite-sample performance of our nonparametric proposal to compute h-step ahead predictions of recession probabilities, $P_{NP}(z_{t+h}=1)$, for $h=1,2,3$. In addition, we use the simulations to evaluate how data problems, such as influential points, structural breaks, heteroskedasticity, and ARCH effects, might affect forecast performance. To facilitate comparisons with Wecker's (1979) linear approach, we also include the forecasts of recession probabilities using an autoregressive model, $P_L(z_{t+h}=1)$.

For the sake of comparison, we have included the forecasts of a Markov-switching autoregressive model of order q, MSAR(q), which is one of the most popular approaches used to compute recession probabilities. Following Hamilton (1989), we assume that the dynamics of y_t are governed by an unobservable regime-switching state variable, s_t. The model can be stated as

$$y_t=\mu_{s_t}+a_1(y_{t-1}-\mu_{s_{t-1}})+\ldots+a_q(y_{t-q}-\mu_{s_{t-q}})+\epsilon_t, \tag{2.13}$$

[4] It implies that $P^{t}_{\pi(\tau)}=P^{t}_{\pi}$ if $\pi(\tau)=\pi$, for all $\pi\in S_{h+1}$.

[5] It is worth pointing out that, since $P_{NP}(z_{t+h}=1)=E(z_{t+h})$, expression (2.3) is an unbiased estimator of the probability of recession.

where $\epsilon_t \sim iidN(0, \sigma^2)$.[6]

Within this framework and assuming that $\mu_0 > \mu_1$, one can label $s_t = 0$ and $s_t = 1$ as the expansion and recession states at time t, respectively. In addition, it is commonly supposed that the state variable evolves following an irreducible 2-state Markov chain whose transition probabilities are defined by

$$p(s_t = j|s_{t-1} = i, s_{t-2} = h, ..., I_{t-1}) = p(s_t = j|s_{t-1} = i) = p_{ij}, \tag{2.14}$$

where $i, j = 0, 1$ and $I_t = \{y_1, \ldots, y_t\}$ is the information set up to period t.

Hamilton (1989) described a forward filter to store the filtered probabilities of recession $P(s_t = 1|\theta, I_t)$, where $\theta = (\mu_0, \mu_1, a_1, \ldots, a_q, p_{00}, p_{11})$ and to provide a maximum likelihood estimation of model parameters $\hat{\theta}$. Using the parameter estimates, it is easy to compute inferences about the business cycle regime at $t + 1$ as

$$P(s_{t+1} = 1|\hat{\theta}, I_t) = \sum_{i=0}^{1} P(s_t = i|\hat{\theta}, I_t)p_{i1}. \tag{2.15}$$

This expression can be used recursively to obtain h-period ahead forecasts $P(s_{t+h} = 1|\hat{\theta}, I_t)$.

We quantify the ability of the Wecker (1979) method, the Markov-switching model, and our nonparametric approach to forecasting the h-step ahead state of the business cycle with the help of three different metrics. The first metric is the Brier score, BS, which is the mean square error of recession probability. A Brier score of 0 means perfect accuracy, and a Brier score of 1 means perfect inaccuracy. For more detailed information this metric is also computed only for recessions (BSR) and expansions (BSE).

The second metric is the area under the receiver operating characteristic curve, namely $AUROC$ (Berje and Jorda, 2011), which is a measure of the overall performance of a binary classifier, it considers the trade-off between the true positive rate and the false positive rate. The $AUROC$ takes values between 0.5 for a random classifier and 1 for a perfect classifier. Regarding this metric, we define the True Posititive rate (TPR) as the probability of predicting recessions that actually become recessions and the True Negative rate (TNR) as the probability of predicting expansions that actually become expansions.

The third metric is Cohen's Kappa coefficient (Cohen, 1960), which is a chance-corrected measure of agreement between the classification of the forecasting techniques and actual recessions. This metric, denoted by κ, reaches 1 for complete agreement. A more detailed explanation of these metrics can be found in Appendix C.

[6]The dynamics can be adapted to account for regime shifts in the autoregressive parameters and in the variance. In addition, the nonlinearities can be imposed in the mean or in the drift of the time series.

2.3.1. Nonparametric model performance

We start the simulations by generating $r = 1, \ldots, R = 500$ business cycle sequences s_t^r of expansions ($s_t^r = 0$) and recessions ($s_t^r = 1$) of length $T = 500$ that follow 2-state Markov chains. To ensure that these dummies share the standard business cycle dynamics, we use the NBER dates to compute the percentage of quarters classified as expansions followed by expansions and the percentage of quarters classified as recessions followed by recessions in the period 1955.2-2022.3. According to this analysis, we set $p_{00} = 0.9$ and $p_{11} = 0.6$.

The data-generating process is an MSAR(1) as outlined in (2.13) with noisy terms $\epsilon_t^2 \sim iidN(0, 0.5)$. Using s_t^r, the dynamics of time series y_t^r are generated by setting the differences of the within-state means $\mu_0 - \mu_1 = 0.5$ and a state-independent autoregressive parameter $a_1 = 0.2$. With this data-generating process, we examine forecasting performance in an out-of-sample scenario.

The results of this exercise for the nonparametric approach are displayed in Table 1. The baseline simulations, whose results are presented in Panel A, show that the accuracy of the probability forecasts deteriorates very little as the forecasting horizon increases, as documented by the Brier score. Similarly, the $AUROC$ metric is around 0.8 regardless of the forecasting horizon, indicating that our nonparametric method presents a good discriminating ability to distinguish the state of the generated business cycles. Finally, Cohen's kappa suggests a reasonable level of agreement between the one-period forecasts of the nonparametric method and the generated business cycles. Nevertheless, κ tends to diminish slightly as the forecasting horizon increases.

Panel B of Table 1 evaluates the effect of the persistence of the generated time series, measured by the autoregressive parameter a_1, on the nonparametric probability forecasts. The results indicate that the forecasting performance is somehow worsened by higher persistence, mainly through BSR. However, the deterioration of the forecasting performance is not dramatic since the persistence does not change this metric substantially when the autoregressive parameter rises from $a_1 = 0.2$ to $a_1 = 0.5$ or $a_1 = 0.8$.

To examine the effect of the sample size in the nonparametric approach, Panel C displays the estimates of the forecasting performance metrics for data-generating processes of sample sizes $T = 250$ and $T = 1000$. In these two cases, the figures are similar to those obtained in the baseline scenario, indicating that the model's performance is invariant to time series of reasonable sample sizes.

We also study the effects of uncertainty on forecasting performance by generating noisy terms with variances of $\sigma^2 = 1$ and $\sigma^2 = 1.5$, shown in Panel D. As expected, the ability of the model to compute inferences on the generated business cycle deteriorates substantially when variance σ^2 increases. Thus, we find that noisy scenarios deteriorate the ability of the

nonparametric procedure to classify the periods into recessions and expansions.

In addition, we evaluate the effects of business cycle phase persistence on forecasting performance by performing simulations with combinations (p_{00}, p_{11}) of $(0.6, 0.6)$ and $(0.9, 0.9)$. The results are shown in Panel E. When a business cycle phase becomes more persistent, the Brier score falls, whereas the overall performance of the binary classifier ($AUROC$) and the overall agreement (κ_h) of the forecasts and the generated cycles increase.

Finally, to examine how the business cycle signal affects forecasting performance, we also set the within-state difference $\mu_0 - \mu_1$ to 1 and 2. Panel F shows that larger differences of within-state means substantially improve the performance of the nonparametric model as the signal-to-noise of the data-generating process increases. Thus, large differences between within-state means facilitate the classification of the time periods into recessions and expansions.

2.3.2. Comparison of the methods' performance under data problems

Despite the good performance of the model in providing statistical inferences of the generated business cycles in the baseline scenario, the data problems that characterize the economic dynamics in empirical applications could lead to potential performance deterioration. To evaluate these potential adverse effects, we conduct an out-of-sample forecasting exercise with outliers, structural breaks, heteroskedasticity, and ARCH dynamics.

In addition, we are interested in evaluating the differences in performance deterioration that the data problems may cause in Markov-switching, linear and nonparametric specifications to forecast recession probabilities. For this purpose, Panels A, B, and C of Table 2 display the Brier scores, $AUROC$, and Kappa coefficients achieved by these three alternative forecasting proposals. To facilitate comparisons, the first row in each panel reports the results of the baseline scenario. The table shows that the three models perform well forecasting business cycle phases. For $h = 1$, the three models display similarly low BS, and high $AUROC$ and κ. Notably, the forecasts that deteriorate the least with the prediction horizon are those of the nonparametric model. Regardless of the statistic, the nonparametric model has the greatest classification ability for $h = 2$ and $h = 3$.

To assess the performance deterioration caused by extremely large observations, we generate additive outliers in the simulated time series that are consistent with the magnitude of the large GDP growth rates observed in 2020. Specifically, we add an additive outlier of -15 standard deviations from the mean of the simulated time series at $t = 100$, followed by a 10 standard deviations outlier in $t + 1$ to the baseline data-generating process. In line with the results reported in Table 2, of the three models, the one that fails the most in performing business cycle inferences due to the large number of observations is the Markov-switching model. Intuition indicates that the estimated mean in the low-growth regime is dominated by the first

outlier and the sample after this date is classified as expansion regardless of the value of the time series. For this reason, BSR tends to 1 while TPR and $AUROC$ are close to 0 and 0.5, respectively.

For one-period forecasting horizons, the performance deterioration of the probability forecasts computed with the linear model is not as significant as that of the Markov-switching model. However, the deterioration is much greater for forecasting horizons larger than 1, as demonstrated by the low values of $AUROC$, which fall to about 0.5, indicating no better classification ability than the toss of a coin strategy. Performance deterioration in forecasting recessions is also evident in the large value of BSR, which is 0.72, and the low value of κ, which tends to 0.

Remarkably, the forecasts computed with the nonparametric algorithm do not show any significant deterioration in the metrics used to examine the performance of the h-step ahead forecasts of the business cycle, regardless of the forecasting horizon considered. As Panel C of Table 2 shows, the metrics are almost invariant when the outlier is introduced into the data-generating process and BS, $AUROC$, and κ are roughly similar to the baseline scenario.

The second robustness check focuses on structural breaks. To examine this data problem, we generate the last three-fourths of the sample as in the baseline scenario ($\mu_0 = 0.25$) while we set $\mu_0 = 1$ for the first fourth of the sample, with the rest of the parameters unaltered. The third row of each panel of Table 2 shows that, as in the case of the outlier, the structural break produces substantial deterioration in the performance of the Markov-switching and linear specifications. However, the changes in the performance of the nonparametric proposal are, again, negligible.

Finally, we evaluate the models' performance to static and serially correlated heteroskedasticity. To simulate data with static heteroskedasticity, we generate the last three-fourths of the sample as in the baseline scenario ($\sigma^2 = 0.5$) while we set $\sigma^2 = 2.5$ for the first fourth of the sample. Serially correlated heteroskedasticity is achieved by generating disturbances $\epsilon_t = \sigma_t u_t$, where u_t is a normalized independent Gaussian process and $\sigma_t = 0.2 + 0.8\epsilon_{t-1}$. Again, the performance deterioration of Markov-switching and linear autoregressive forecasts is substantial, although a bit less severe than in the cases of outliers and structural breaks. As in the previous two scenarios, the deterioration in the performance of the linear and Markov-switching approaches is much greater than in our nonparametric proposal.

2.4 — Empirical example

In this section, we assess the empirical reliability of Markov-switching specifications, linear autoregressive models, and our nonparametric approach to provide accurate in-sample forecasts in the run-up to business cycle recessions in the G7 countries. In addition, we examine the impact of the COVID-19 recession, which led to record falls and recoveries, on forecasting

performance. For this purpose, we develop the analysis with a sample that ends in 2019 and a complete sample that includes the COVID-19 pandemic data ending in the third quarter of 2022.

The COVID-19 recession is the latest in our data sets. Thus, we are precluded from examining its effect in out-of-sample exercises as in the section devoted to Monte Carlo simulations. To assess the effect that the observations recorded in 2020 will have on the future performance of the forecasting approaches, we focus on the impact of these influential observations on historical business cycle dating. Notice that we do not pursue ad-hoc shortcuts as temporary solutions to this problem, such as using additional regimes or shortening the sample used to estimate model parameters.

We use data sets of the growth rates of seasonally adjusted real GDP for the Group of Seven (G7) countries (USA, UK, Germany, France, Italy, Canada, and Japan). The data come from the OECD Main Economic Indicators. The sample starts in 1955.2 for the UK and US, 1960.2 for Italy, Canada, and Japan, 1964.1 for Germany, and 1969.1 for France. It ends in 2022.3 for all the countries. To evaluate the empirical performance of the forecasting approaches, we use the reference cycle dates provided by the dating committees of the NBER in the case of the US and the Economic Cycle Research Institute (ECRI) for the rest of the countries.

2.4.1. Pre-COVID-19 data

In the first approximation to the analysis of forecasting performance, we focus on the historical ability of the three forecasting approaches to classify the dates into expansion and recession in constrained samples that end in the last quarter of 2019. Remarkably, the results reported in Panel C of Table 3 show that the probability forecast of our nonparametric procedure is in close agreement with the reference cycles for all the countries. Regardless of the forecasting horizon, our approach results in a low Brier score, an $AUROC$ much higher than 0.5, and a large kappa coefficient, all of which are comparable to those reached by the Markov-switching model (Panel A) and the linear specification (Panel B).

In terms of ROC curves, the Markov-switching model achieves the best business cycle performance across all the possible classification thresholds for all the countries but France, Germany, and Italy, in part because their higher volatilities tend to diminish the ability of the Markov-switching model to separate the states.[7] Although this approach shows high values for the US, UK, and Canada, the magnitudes reported for the linear and the nonparametric approaches are also significantly greater than 0.5, indicating their considerable discriminating ability. The apparently better performance of the Markov-switching approach relies on its ability to classify expansions because this advantage vanishes when examining the ability to

[7]This result agrees with those of Carstensen et al.(2020).

correctly identify recessions with TPR statistics.

Measured by the Brier score, the differences in forecasting performance across the three competing approaches are minor in all the cases. The nonparametric model slightly outperforms in one and two-quarter forecasts, and the linear approach outperforms in three-quarter forecasts. Regarding the kappa index, the data reveal uniformly closer agreement between the nonparametric probabilities of recession the official recessions than with the other two forecasting approaches. The numbers reported in Table 3 show that the kappa coefficients of the nonparametric approach are substantially larger than those of the other two competitors, regardless of the forecasting horizon and the country of the sample.

To illustrate the good performance of the three dating processes with samples that do not include the pandemic period, Figure 2.1 displays the growth rates of quarterly real GDP for the US from 1955.2 to 2019.4 (Panel A) and the 2-quarter ahead predictions of the probabilities of recession obtained from a Markov-switching model (Panel B), a linear specification (Panel C) and our nonparametric approach (Panel D), which are obtained using full-sample parameter estimates. To facilitate comparisons, the panels include the dates of economic recessions as determined by the NBER, which are shaded.

The Markov-switching model estimates within-expansion and within-recession means of $\hat{\mu}_0 = 0.91$ and $\hat{\mu}_1 = -0.56$, respectively. Panel B of Figure 2.1 plots the two-horizon forecast recession probabilities $P(s_{t+2} = 1|\hat{\theta}, I_t)$, with $\hat{\theta}$ estimated using data up to 2019.4. These probabilities produce satisfactory data classification into expansions and recessions, reproducing the NBER chronology very closely. During periods that the NBER classifies as expansions, the probabilities of recession are usually close to zero. At the NBER peaks, the probabilities rise above 30% and remain at these levels until the NBER troughs.

Moving to Wecker's approach, we fit an AR(2) model to the GDP growth rates, whose parameters are also estimated using data up to 2019.4. This specification produces estimates of the autoregressive parameters of $\hat{a}_1 = 0.26$ and $\hat{a}_2 = 0.13$, respectively, which are statistically significant. At each period, we use (2.3) to compute the 2-period forecasts of recession probability, $P_L(z_{t+2} = 1)$, which are plotted in Panel C of Figure 2.1. The predicted probabilities of recession also align well with the official NBER business cycle, although to a lesser degree than the Markov-switching specification. Generally, when the probability of recession exceeds 40%, a recession follows, while recession probabilities fall below 40% in recession troughs.

Finally, Panel D of Figure 2.1 plots the 2-quarter forecasts of recession probabilities computed as expression (2.12). Despite the simplicity of computing the forecasts, the figure shows that the new, nonparametric assessment of recession probabilities is proficient at capturing the NBER-referenced business cycle chronology. Specifically, the forecasts of recession probabilities always jump to almost one at the point of NBER recessions (shown in shaded areas) and remain much lower at NBER expansions.

2.4.2. Complete data set

Despite the good performance of the three alternative methods when the samples end in 2019, the COVID-19 pandemic in 2020 hit the national economies worldwide with unprecedented force causing record fluctuations in GDP growth that have substantially altered the findings obtained with pre-pandemic data. Particularly, the G7 countries faced one of their sharpest declines during the health restrictions established in the first quarter of 2020, followed by exceptionally rapid rebounds when the restrictions were relaxed and stimulus measures came into effect.

In Table 4, we evaluate the impact of the extraordinaryly high growth rates reported in 2020 on forecasting performance by extending the sample to the third quarter of 2022. Regardless of the country, the table shows that the Markov-switching specification has lost its business cycle classification abilities. For each country, the model identifies only one recession in 2020 while classifying the rest of the sample dates as expansions. This implies that BSR tends to be one, and BSE tends to be zero. For this model, the $AUROC$ metrics are close to 0.5, suggesting no better performance than a random classifier. In addition, the kappa coefficients are always close to zero.

The deterioration suffered in the linear autoregressive forecasts is a bit less severe than in the Markov-switching approach. However, we find that BSR increases substantially and the TPR falls dramatically, as does κ_h, indicating that this model is not very useful for anticipating future recessions. In addition, Table 4 reveals that the losses in forecasting accuracy when pandemic data is included increase as the forecasting horizon expands.

In contrast to the Markov-switching model and the linear autoregressive specification, Table 4 shows that our nonparametric approach is scarcely affected by the 2020 data. Regardless of the country, when the extreme data are included, the model still agrees closely with the reference cycles, has good overall performance as a binary classifier, and is consistent with the turning points determined by the national dating committees.

To illustrate this graphically, we enlarge the sample in Figure 2.1 with US GDP growth rates observed up to 2022.3. Panel A of Figure 2.2 shows that the US suffered from the deepest recession and the most significant recovery on record in two consecutive quarters (-8.9% and 7.5% in 2020.2 and 2020.3, respectively). These two extreme observations blur the patterns of the time series dynamics observed in the pre-pandemic data. A straightforward explanation is that the two influential observations drastically skewed the empirical distribution of the GDP.

In the case of the Markov-switching specification, the two influential observations in 2020 affect the estimate of the within-recession mean dramatically. It falls to $\hat{\mu}_1 = -9.07$, leaving the within-recession mean almost unaltered at $\hat{\mu}_0 = 0.77$. The resulting 2-period ahead forecasts of recession probability, displayed in Panel B of Figure 2.2 show that the cyclical interpretation of

high and low growth states fails spectacularly. The substantial drop in GDP growth documented in 2020.2, which is identified as a low-growth state, is so influential that the model relegates all the previous recessions to a high-growth state. This invalidates the standard Markov-switching approach as a tool for anticipating future recessions unless we use ad-hoc shortcuts such as allowing changes in model parameters in 2020 or estimating model parameters using data only up to 2019.

Regarding Wecker's technique, it is well-known that extreme observations greatly affect the forecasting performance of linear autoregressive models and tend to bias the full-sample parameter estimates. Fitting an AR(2) to the enlarged sample of growth rates produces autoregressive parameters that fall dramatically to $\hat{a}_1 = 0.03$ and $\hat{a}_2 = 0.08$ and become statistically non-significant. We use these estimates to simulate the forecast paths and the recessionary indicator required to compute the two-period forecasts of recession probabilities. Panel C of Figure 2.2, which displays these probabilities, shows that Wecker's technique also fails to predict the NBER-referenced recessions when the two extreme observations released in 2020 are used to compute the model-based forecasts.

Notably, Panel D of Figure 2.2 shows that enlarging the sample with data up to 2022.3 does not deteriorate the performance of the new nonparametric approach. In this case, estimating the model parameters from samples that use the extreme observations reported in 2020 leads to recession probabilities that continue to be remarkably similar to NBER business cycle dating. This shows that, unlike the cases of autoregressive models and Markov-switching specifications, our proposal minimizes the importance of extreme observations in the time series used to perform the forecasts of recession probabilities.

To sum up, the two influential observations of GDP growth rates in 2020 have (and will have in the future) devastating effects on business cycle identification from standard parametric models and calls into question whether they will be useful for dating business cycles from the COVID-19 period on. We show that our nonparametric proposal is robust to this and other data-generating problems, such as structural breaks and when errors are heteroscedastic or present ARCH dynamics[8].

2.5 — Conclusions

Providing economic agents with early warning systems that give advanced notice of future business cycle developments has become an issue of great interest in economics. For this purpose, some parametric approaches, such as the Markov-switching model advocated by Hamilton (1989) or the autoregressive forecasts used by Wecker (1979), have become very popular methods to provide recession probabilities. However, the unprecedented magnitude of the recession

[8]All data and codes are available at Github

caused by the recent global pandemic has shown that these tools fail to produce robust dating of business cycle turning points.

To contribute to this literature, this chapter develops a nonparametric extension of the autoregressive forecast method, combined with symbolic dynamics, to compute robust inferences of reference cycle turning points. Our simulations suggest that the method performs well in recession predictions and that, unlike the parametric alternatives, it is robust against outliers, structural breaks, heteroscedasticity, and ARCH effects. This is desirable in early warning systems because, in practice, these data problems are the norm rather than the exception.

Using pre-COVID-19 data, the empirical evidence shows that the historical ability of the nonparametric approach to forecasting the business cycle phases of the G7 countries is similar to that of its parametric competitors. However, when the sample is enlarged with COVID-data, the nonparametric approach substantially outperforms the parametric forecasts, whose performance is hardly better than a random classifier.

We look forward to carrying out future work addressing the following issues. First, we see a natural extension of our approach to developing early warning systems for determined events in many other situations by defining the event under consideration differently. For example, further application areas are forecasting critical transitions in temperature regimes, detecting product defects in manufacturing, or providing statistical classifications of diseases. Second, our empirical application relies on an in-sample approach because the post-COVID-19 data have only been available for roughly two years. As new vintages become available, real-time reassessments will provide new insights into the model's forecasting performance. Third, we see the possibility of adjusting the methodology to a multivariate approach, to take several series into account when predicting recessions. While the number of symbols would largely grow with the increase of series, the use of tools from the symbolic literature such as transcripts to achieve dimensionality reduction could help in pursuing this extension.

2.6 — References

Abd Alla, F. 2004. Distributions of order patterns in time series. PhD dissertation, Greifswald.

Anderson H., Teräsvirta T., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregression models. *Journal of Applied Econometrics* 7: S199-136.

Antolin-Diaz, J., Drechsel, T., and Petrella, I. 2021. Advances in nowcasting economic activity: Secular trends, large shocks and new data. CEPR Discussion Papers 15926.

Berge, T.J. and Jordà, O. 2011. Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics* Vol 3, N0. 2, 246-277.

Camacho, M., and Perez Quirós, G. 2002. This is what the leading indicator lead. *Journal of Applied Econometrics* 17: 61-80.

Camacho, M., Romeu, A., and Ruiz, M. 2021. Symbolic transfer entropy test for causality in longitudinal data. *Economic Modelling* 94: 646-661.

Carriero, A., Clark, T., Marcellino, M., and Mertens, E. 2021. Measuring uncertainty and its effects in the COVID-19 era. CEPR Discussion Papers 15965.

Carstensen,K., Heinrich, M., Reif, M., and Wolters, M.H. 2020. Predicting ordinary and severe recessions with a three-state Markov-switching dynamic factor model. *International Journal of Forecasting* 37 (3), 1328-1328.

Chauvet, M., and Hamilton, J. 2006. Dating Business Cycle Turning Points. In Milas, C., Rothman, P., van Dijk, D., and Wildasin, E. (eds) *Nonlinear Time Series Analysis of Business Cycles*. Emerald Group Publishing Limited: Bingley.

Chauvet, M., and Piger J. 2008. A comparison of the real-time performance of business cycle dating methods. *Journal of Business and Economic Statistics* 26: 42-49.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20: 37–46.

Collet, P., and Eckmann, J. 2009. Iterated maps on the interval as dynamical systems. Birkhauser, Basel.

Dore, M., Matilla, M. and Ruiz, M. 2014. A permutation entropy based test for causality: The volume-stock price relation. *PHYSICA A* 398: 280-288.

Estrella, A. and Mishkin, F. 1998. Predicting US recessions: financial variables as leading indicator. *Review of Economic and Statistics* 80: 45-61.

Hamilton, J., and Perez-Quiros, G. 1996. What do the leading indicators lead? *Journal of Business* 69: 27-49.

Hamilton, J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle *Econometrica* 57: 357-384.

Hou, Y., Liu, F., Gao, J., Cheng, Ch., and Song, Ch. 2017. Characterizing complexity changes in Chinese stock markets by permutation entropy. *Entropy* 19: 514.

Leiva-Leon, D., Perez-Quiros, G., and Rots, E. 2020. Real-time weakness of the global economy. European Central Bank Working Paper No 2381.

McGrane, M. 2022. A Markov-switching model of the unemployment rate. Congressional Budget Office No. 2022-05.

Ng, S. 2021. Modeling macroeconomic variations after COVID-19. NBER Working Papers No.

29060.

Lenza, M., and Primiceri, G. 2020. How to estimate a VAR after March 2020. NBER Working Papers No. 27771.

Tino P., Schittenkopf, Ch., Dorffner, G., and Dockner, E. 2000. A symbolic dynamics approach to volatility prediction. In Abu-Mostafa, Y., LeBaron, B., Lo, A. and Weigend, A. (eds) *Computational Finance*. MIT Press: Cambridge.

Wecker, W. 1979. Predicting the turning points of a time series. *Journal of Business* 52: 35-50.

2.7 — Appendix A

Let y_t be a time series that follows the following autoregressive process of order 2

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t, \tag{A.1}$$

where ϵ_t is a white noise Gaussian disturbance term with a mean of 0 and variance σ^2. For a one-step forecasting horizon $h = 1$, simulations of y_{t+1} can be developed from the distribution

$$y_{t+1} \sim N(c + a_1 y_t + a_2 y_{t-1}, \sigma_2). \tag{A.2}$$

When the forecasting horizon is $h = 2$, simulations of $(y_{t+2}, y_{t+1})'$ can be performed by

$$\begin{pmatrix} y_{t+2} \\ y_{t+1} \end{pmatrix} \sim N\left(\begin{pmatrix} (1+a_1)c + (a_1^2 + a_2^2)y_t + a_1 a_2 y_{t-1} \\ c + a_1 y_t + a_2 y_{t-1} \end{pmatrix}, \sigma^2 \begin{pmatrix} (1+a_1^2) & a_1 \\ a_1 & 1 \end{pmatrix}\right). \tag{A.3}$$

Finally, when $h = 3$, simulations of $(y_{t+3}, y_{t+2}, y_{t+1})'$ can be obtained from

$$\begin{pmatrix} y_{t+3} \\ y_{t+2} \\ y_{t+1} \end{pmatrix} \sim N\left(\begin{pmatrix} (1 + a_1 + a_2 + a_1^2)c + (a_1^3 + 2a_1 a_2)y_t + (a_1^2 a_2 + a_2^2)y_{t-1} \\ (1+a_1)c + (a_1^2 + a_2^2)y_t + a_1 a_2 y_{t-1} \\ c + a_1 y_t + a_2 y_{t-1} \end{pmatrix}, Q\right), \tag{A.4}$$

where

$$Q = \sigma^2 \begin{pmatrix} (a_1^2 + a_2)^2 + a_1^2 + 1 & a_1(a_1^2 + a_2) + a_1 & a_1^2 + a_2 \\ a_1(a_1^2 + a_2) + a_1 & (1 + a_1^2) & a_1 \\ a_1^2 + a_2 & a_1 & 1 \end{pmatrix}. \tag{A.5}$$

2.8 — Appendix B

Let $\{y_t\}_{t=1}^{T}$ be a stationary Gaussian process and ρ_i^t its autorrelation of order i computed with information up to t. In addition, let $h = 1, 2, 3$ be the forecasting horizon, let $Y_{h+1}(\tau)$ be the $(h+1)$-dimensional history of this time series starting at τ, and let $\pi \in S_{h+1}$ be its corresponding symbol. Finally, let P_π^t be the forecast probability of that symbol. For a forecasting horizon $h = 1$, the probability of the symbols that belong to S_2 are

$$P_{(0,1)}^t = P_{(1,0)}^t = \frac{1}{2}. \tag{B.1}$$

For a forecasting horizon $h = 2$, the probability of the symbols that belong to S_3 become

$$P_{(0,1,2)}^t = P_{(2,1,0)}^t = \frac{1}{\pi} arcsin\left(\frac{1}{2}\sqrt{\frac{1-\rho_2^t}{1-\rho_1^t}}\right) \tag{B.2}$$

and

$$P_{(0,2,1)}^t = P_{(1,2,0)}^t = P_{(2,0,1)}^t = P_{(1,0,2)}^t = \frac{1}{4}\left(1 - \frac{2}{\pi} arcsin\left(\frac{1}{2}\sqrt{\frac{1-\rho_2^t}{1-\rho_1^t}}\right)\right) \tag{B.3}$$

Finally, for a forecasting horizon $h = 3$, the probability of the symbols that belong to S_4 are

$$P_{(0,1,2,3)}^t = P_{(3,2,1,0)}^t = \frac{1}{8}\left(1 + \frac{2}{\pi}\left(arcsin(\alpha_1^t) + 2arcsin(\alpha_2^t)\right)\right), \tag{B.4}$$

$$P_{(2,0,3,1)}^t = P_{(1,3,0,2)}^t = \frac{1}{8}\left(1 + \frac{2}{\pi}\left(2arcsin(\alpha_3^t) + arcsin(\alpha_4^t)\right)\right), \tag{B.5}$$

$$P_{(3,1,2,0)}^t = P_{(0,2,1,3)}^t = \frac{1}{8}\left(1 + \frac{2}{\pi}\left(arcsin(\alpha_4^t) - 2arcsin(\alpha_5^t)\right)\right). \tag{B.6}$$

$$P_{(1,0,3,2)}^t = P_{(2,3,0,1)}^t = \frac{1}{8}\left(1 + \frac{2}{\pi}\left(2arcsin(\alpha_6^t) + arcsin(\alpha_1^t)\right)\right), \tag{B.7}$$

$$P_{(0,1,3,2)}^t = P_{(1,0,2,3)}^t = P_{(2,3,1,0)}^t = P_{(3,2,0,1)}^t = \frac{1}{8}\left(1 + \frac{2}{\pi}\left(arcsin(\alpha_7^t) - arcsin(\alpha_1^t) - arcsin(\alpha_5^t)\right)\right), \tag{B.8}$$

$$P_{(2,0,1,3)}^t = P_{(0,2,3,1)}^t = P_{(3,1,0,2)}^t = P_{(1,3,2,0)}^t = \frac{1}{8}\left(1 + \frac{2}{\pi}\left(arcsin(\alpha_7^t) - arcsin(\alpha_4^t) - arcsin(\alpha_5^t)\right)\right), \tag{B.9}$$

$$P_{(0,3,1,2)}^t = P_{(3,0,2,1)}^t = P_{(2,1,3,0)}^t = P_{(1,2,0,3)}^t = \frac{1}{8}\left(1 + \frac{2}{\pi}\left(arcsin(\alpha_3^t) + arcsin(\alpha_8^t) - arcsin(\alpha_5^t)\right)\right), \tag{B.10}$$

and

$$P_{(0,3,2,1)}^t = P_{(3,0,1,2)}^t = P_{(1,2,3,0)}^t = P_{(2,1,0,3)}^t = \frac{1}{8}\left(1 + \frac{2}{\pi}\left(arcsin(\alpha_6^t) - arcsin(\alpha_8^t) + arcsin(\alpha_2^t)\right)\right), \tag{B.11}$$

where

$$\alpha_1^t = \frac{2\rho_2^t - \rho_1^t - \rho_3^t}{2(1-\rho_1^t)}, \qquad \alpha_2^t = \frac{2\rho_1^t - \rho_2^t - 1}{2(1-\rho_1^t)}, \qquad \alpha_3^t = \frac{\rho_2^t - \rho_1^t + \rho_3^t - 1}{2\sqrt{(1-\rho_2^t)(1-\rho_3^t)}},$$

$$\alpha_4^t = \frac{\rho_1^t - \rho_3^t}{2(1-\rho_2^t)}, \qquad \alpha_5^t = \frac{1}{2}\sqrt{\frac{1-\rho_2^t}{1-\rho_1^t}}, \qquad \alpha_6^t = \frac{\rho_1^t + \rho_3^t - \rho_2^t - 1}{2\sqrt{(1-\rho_1^t)(1-\rho_2^t)}},$$

$$\alpha_7^t = \frac{\rho_1^t + \rho_2^t - \rho_3^t - 1}{2\sqrt{(1-\rho_1^t)(1-\rho_2^t)}}, \qquad \alpha_8^t = \frac{\rho_1^t - \rho_2^t}{\sqrt{(1-\rho_1^t)(1-\rho_3^t)}}.$$

2.9 — Appendix C

The Brier score is computed as the average over the R simulations of the squared deviations of the h-period recession probability forecasts, $\hat{P}_h^{tr}$, from a binary value, s_{t+h}^r, which takes the value of one in actual recessions:

$$BS_h = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{T}\sum_{t=1}^{T}(\hat{P}_h^{tr} - s_{t+h}^r)^2, \tag{C.1}$$

where $\hat{P}_h^{tr}$ is $P_{NP}(z_{t+h} = 1)$ in the case of the nonparametric forecasts proposed in this chapter. It is $P_L(z_{t+h} = 1)$ in the case of the autoregressive forecasts, and it is $P(s_{t+h} = 1|\hat{\theta}^r, I_t)$ in the case of the Markov-switching autoregressive forecasts for $h = 1, 2, 3$, and $r = 1, \ldots, R$. A Brier score of 0 means perfect accuracy, and a Brier score of 1 means perfect inaccuracy. When the measure focuses only on those time periods where a recession occurs ($s_t^r = 1$), we call it the Brier Score of Recessions (BSR_h), whereas when it focuses on expansions ($s_t^r = 0$), we call it the Brier Score of Expansions (BSE_h).

The second metric used to measure the models' performance follows the lines suggested by Berje and Jorda (2011). In particular, we measure the recession/expansion classification ability of the three forecasting methods using the Receiver Operating Characteristic (ROC) framework. In particular, given a threshold c, a recession is called when $\hat{P}_h^{tr} > c$, whereas an expansion is called when $\hat{P}_h^{tr} \leq c$. For each simulation, we can define the True Positive rate, $TP_h^r(c)$, and the False Positive rate, $FP_h^r(c)$, as

$$TP_h^r(c) = P(\hat{P}_h^{tr} > c|s_{t+h}^r = 1), \tag{C.2}$$

$$FP_h^r(c) = P(\hat{P}_h^{tr} > c|s_{t+h}^r = 0). \tag{C.3}$$

This implies that $TP_h^r(c)$ is the probability of calling recessions when there are actually recessions, and $FP_h^r(c)$ is the probability of calling recessions when there are actually expansions. We refer to the True Positive Rate, TPR_h, as the average of $TP_h^r(c)$ for all $r = 1, 2, \ldots, R$, and all c from 0 to 1 at steps of 0.001,

$$TPR_h = \frac{1}{1000R}\sum_{r=1}^{R}\sum_{c=0.001}^{1} TP_h^r(c). \tag{C.4}$$

Similarly, we define the True Negative rate, TNR_h as

$$TNR_h = \frac{1}{1000R}\sum_{r=1}^{R}\sum_{c=0.001}^{1}(1 - FP_h^r(c)). \tag{C.5}$$

The ROC curve represents the trade-off between $TP_h^r(c)$ and $FP_h^r(c)$ for different thresh-

olds c. Concretely the *ROC* curve is represented plotting the points $(FP_h^r(c), TP_h^r(c))$ on a $[0,1]\times[0,1]$ plane for all possible thresholds c. When $\hat{P}_h^{tr}$ is an uninformative classifier with respect to the phase cycle, the *ROC* curve coincides with the main diagonal line, and, when it is a perfect classifier the *ROC* curve is on the upper left part of the unit quadrant. A standard measure of overall classification ability is the area under the *ROC* curve, denoted by $AUROC_h^r$. This quantity takes values between 0.5 for a random classifier and 1 for a perfect classifier (see Berge and Jorda, 2011). Then, we take

$$AUROC_h = \frac{1}{R}\sum_{r=1}^{R} AUROC_h^r, \tag{C.6}$$

which measures the business cycle classification ability of the h-step forecasting probabilities.

Finally, we use Cohen's kappa coefficient, originally developed by Cohen (1960). It is a chance-corrected measure of agreement between forecasting technique classifications and real recessions. For a given threshold c, the kappa coefficient is calculated as

$$\kappa_h^r(c) = \frac{P_a^r(c) - P_r^r(c)}{1 - P_u^r(c)}, \tag{C.7}$$

where P_a^r denotes the probability of overall agreement and P_u^r is the probability of hypothetical probability of chance agreement for the r-th simulation. Specifically, these probabilities are calculated as:

$$P_a^r(c) = \frac{1}{T-h}\sum_{t=1}^{T-h}\left[I(\hat{P}_h^{tr} > c)s_{t+h}^r + I(\hat{P}_h^{tr} < c)(1 - s_{t+h}^r)\right], \tag{C.8}$$

and

$$P_u^r(c) = \frac{1}{T-h}\sum_{t=1}^{T-h} I(\hat{P}_h^{tr} > c)\frac{1}{T-h}\sum_{t=1}^{T-h} s_{t+h}^r + \frac{1}{T-h}\sum_{t=1}^{T-h} I(\hat{P}_h^{tr} < c)\frac{1}{T-h}\sum_{t=1}^{T-h}(1 - s_{t+h}^r) \tag{C.9}$$

where $I(\cdot)$ is an indicator function taking the value 1 for a true statement.

Using this notation, the calculation of Cohen's metric, κ_h^r, for a particular forecasting horizon h, is given by the average of all kappa coefficients:

$$\kappa_h = \frac{1}{1000R}\sum_{r=1}^{R}\sum_{c=0.001}^{1}\kappa_h^r(c) \tag{C.10}$$

The interpretation of a given magnitude of κ_h is somehow problematic.[9] Nonetheless, it is straightforward to check that Cohen's κ is equal to 1 when there is complete agreement between the two classifiers while it is equal to 0 when there is no agreement between the classifiers other

[9]Cohen's κ tends to increase when recessions and expansions are equiprobable and when expansions and recessions are distributed asymmetrically by the two classifiers.

than what would be expected by chance. The metric can be negative when the agreement between the two classifiers is worse than random.

2.10 — Tables

Table 1. Nonparametric model's performance

	h=1							h=2							h=3						
	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ
Panel A. Baseline model: $a_1 = 0.2$; $T = 500$; $\sigma^2 = 0.5$; $p_{00} = 0.9$; $p_{11} = 0.6$; $\mu_0 - \mu_1 = 0.5$																					
	0.16	0.36	0.11	0.75	0.52	0.83	0.31	0.13	0.33	0.09	0.80	0.48	0.78	0.23	0.14	0.30	0.10	0.80	0.50	0.74	0.21
Panel B. Changes in time series' inertia																					
$a_1 = 0.5$	0.15	0.42	0.08	0.75	0.49	0.89	0.37	0.12	0.39	0.06	0.81	0.45	0.85	0.29	0.13	0.32	0.08	0.81	0.49	0.77	0.24
$a_1 = 0.8$	0.18	0.65	0.06	0.64	0.31	0.93	0.27	0.16	0.63	0.04	0.74	0.28	0.92	0.23	0.14	0.51	0.05	0.74	0.35	0.87	0.21
Panel C. Changes in sample size																					
$T = 250$	0.17	0.40	0.11	0.73	0.49	0.83	0.28	0.14	0.34	0.09	0.78	0.47	0.78	0.21	0.14	0.32	0.10	0.78	0.48	0.74	0.19
$T = 1000$	0.16	0.36	0.11	0.76	0.52	0.84	0.33	0.13	0.33	0.08	0.80	0.48	0.78	0.24	0.14	0.30	0.10	0.80	0.50	0.75	0.22
Panel D. Changes in variance																					
$\sigma^2 = 1$	0.23	0.44	0.18	0.64	0.45	0.75	0.15	0.18	0.39	0.14	0.67	0.43	0.71	0.11	0.19	0.36	0.15	0.67	0.45	0.68	0.11
$\sigma^2 = 1.5$	0.26	0.47	0.21	0.60	0.42	0.71	0.10	0.20	0.41	0.15	0.62	0.41	0.69	0.08	0.21	0.38	0.16	0.62	0.43	0.66	0.07
Panel E. Changes in states' inertia																					
p_{00}=p_{11}=0.6	0.25	0.36	0.14	0.73	0.52	0.79	0.31	0.22	0.33	0.10	0.76	0.48	0.74	0.23	0.21	0.30	0.12	0.76	0.50	0.71	0.21
p_{00}=p_{11}=0.9	0.21	0.31	0.12	0.78	0.58	0.83	0.40	0.19	0.29	0.09	0.83	0.52	0.78	0.30	0.18	0.26	0.10	0.83	0.54	0.74	0.28
Panel F. Changes in differences of within-state means																					
μ_0-μ_1=1	0.08	0.23	0.04	0.89	0.65	0.94	0.58	0.07	0.24	0.04	0.94	0.57	0.87	0.43	0.09	0.21	0.05	0.94	0.59	0.82	0.37
μ_0-μ_1=2	0.02	0.10	0.01	0.96	0.79	0.99	0.81	0.03	0.15	0.01	0.99	0.68	0.94	0.64	0.04	0.13	0.02	0.99	0.67	0.87	0.54

Notes. For each forecasting horizon $h = 1, 2, 3$, the table shows the total Brier Score (BS), the Brier Score conditional to recessions (BSR) and to expansions (BSE), the area under the ROC curve $AUROC$, the average True Positive Rate (TPR), the average True Negative Rate (TNR) and the average Cohen's kappa coefficient (κ), for the nonparametric approach. Panel A reports the results of the baseline specification and Panels B to F refer to the extent to which departures from the baseline model affect the classification measures.

Table 2. Comparison of the three different approaches under data problems

	h=1							h=2							h=3						
	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ
Panel A. Markov-switching model																					
Baseline	0.28	0.47	0.28	0.66	0.47	0.64	0.10	0.28	0.42	0.25	0.63	0.44	0.63	0.06	0.29	0.44	0.24	0.61	0.43	0.63	0.05
Outliers	0.23	0.75	0.11	0.55	0.19	0.85	0.05	0.23	0.75	0.10	0.55	0.19	0.84	0.03	0.23	0.75	0.10	0.55	0.18	0.84	0.03
Breaks	0.42	0.09	0.50	0.64	0.79	0.38	0.09	0.39	0.14	0.46	0.61	0.71	0.40	0.07	0.37	0.18	0.42	0.59	0.66	0.42	0.05
Hetero	0.41	0.41	0.41	0.52	0.51	0.51	0.01	0.40	0.41	0.40	0.51	0.50	0.51	0.01	0.40	0.41	0.40	0.51	0.50	0.51	0.01
ARCH	0.30	0.66	0.22	0.52	0.28	0.72	0.00	0.30	0.66	0.22	0.51	0.28	0.71	0.00	0.30	0.66	0.22	0.51	0.27	0.71	-0.01
Panel B. Linear model																					
Baseline	0.14	0.51	0.04	0.76	0.32	0.90	0.20	0.15	0.69	0.02	0.75	0.17	0.88	0.05	0.17	0.78	0.01	0.61	0.12	0.90	0.01
Outliers	0.15	0.55	0.04	0.73	0.28	0.90	0.16	0.17	0.72	0.02	0.54	0.16	0.86	0.01	0.17	0.72	0.02	0.52	0.15	0.85	0.00
Breaks	0.14	0.64	0.01	0.73	0.23	0.97	0.23	0.17	0.81	0.01	0.79	0.10	0.97	0.07	0.18	0.87	0.01	0.69	0.07	0.96	0.03
Hetero	0.17	0.51	0.07	0.68	0.32	0.85	0.14	0.16	0.58	0.06	0.56	0.24	0.77	0.01	0.17	0.62	0.05	0.51	0.22	0.79	0.00
ARCH	0.16	0.53	0.06	0.69	0.32	0.87	0.15	0.15	0.64	0.03	0.65	0.21	0.83	0.03	0.16	0.70	0.03	0.57	0.17	0.84	0.01
Panel C. Nonparametric specification																					
Baseline	0.16	0.36	0.11	0.75	0.52	0.83	0.31	0.13	0.33	0.09	0.80	0.48	0.78	0.23	0.14	0.30	0.10	0.80	0.50	0.74	0.21
Outliers	0.16	0.36	0.12	0.75	0.52	0.83	0.31	0.14	0.31	0.10	0.80	0.50	0.76	0.23	0.14	0.31	0.10	0.79	0.50	0.76	0.22
Breaks	0.14	0.41	0.08	0.76	0.47	0.89	0.35	0.12	0.38	0.06	0.84	0.43	0.84	0.25	0.12	0.31	0.08	0.84	0.48	0.79	0.22
Hetero	0.18	0.43	0.12	0.69	0.43	0.81	0.19	0.16	0.39	0.11	0.72	0.40	0.72	0.10	0.17	0.35	0.12	0.72	0.43	0.68	0.09
ARCH	0.20	0.39	0.15	0.70	0.50	0.79	0.24	0.17	0.36	0.12	0.73	0.46	0.74	0.18	0.17	0.32	0.13	0.73	0.48	0.71	0.16

Notes. For the Markov switching, linear and nonparametric approaches, and each forecasting horizon $h = 1, 2, 3$, the table shows the total Brier Score (BS), the Brier Score conditional to recessions (BSR) and to expansions (BSE), the area under the ROC curve $AUROC$, the average True Positive Rate (TPR), the average True Negative Rate (TNR) and the average Cohen's kappa coefficient (κ).

Table 3. Empirical performance in G7 countries: 1955.2-2019.4

	h=1							h=2							h=3						
	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ
Panel A. Markov-switching model																					
US	0.08	0.53	0.01	0.92	0.31	0.93	0.26	0.11	0.70	0.02	0.81	0.17	0.90	0.07	0.11	0.77	0.01	0.69	0.13	0.89	0.02
UK	0.07	0.62	0.01	0.90	0.24	0.96	0.24	0.09	0.74	0.01	0.84	0.15	0.95	0.12	0.10	0.80	0.01	0.76	0.11	0.85	0.07
Canada	0.06	0.51	0.01	0.86	0.36	0.96	0.37	0.07	0.63	0.01	0.81	0.24	0.94	0.21	0.08	0.73	0.01	0.76	0.16	0.93	0.10
France	0.45	0.09	0.54	0.72	0.83	0.35	0.11	0.44	0.10	0.52	0.65	0.80	0.35	0.08	0.43	0.10	0.51	0.58	0.76	0.34	0.06
Germany	0.25	0.97	0.01	0.71	0.02	0.98	-0.01	0.25	0.96	0.01	0.66	0.02	0.98	0.00	0.25	0.96	0.01	0.58	0.02	0.98	0.00
Italy	0.32	0.43	0.29	0.60	0.48	0.62	0.09	0.33	0.47	0.29	0.56	0.46	0.61	0.06	0.34	0.48	0.29	0.52	0.44	0.60	0.04
Japan	0.47	0.01	0.61	0.83	0.95	0.31	0.15	0.46	0.02	0.60	0.76	0.92	0.30	0.12	0.46	0.03	0.59	0.73	0.89	0.29	0.11
Panel B. Linear model																					
US	0.17	0.19	0.17	0.85	0.71	0.74	0.31	0.13	0.29	0.11	0.83	0.50	0.73	0.17	0.13	0.45	0.08	0.68	0.34	0.73	0.05
UK	0.14	0.29	0.12	0.81	0.53	0.78	0.20	0.11	0.42	0.07	0.76	0.37	0.76	0.11	0.11	0.48	0.06	0.73	0.32	0.76	0.07
Canada	0.19	0.21	0.19	0.82	0.70	0.72	0.25	0.12	0.31	0.10	0.76	0.49	0.72	0.16	0.12	0.44	0.08	0.71	0.34	0.73	0.05
France	0.16	0.56	0.07	0.65	0.28	0.83	0.07	0.16	0.69	0.05	0.66	0.17	0.80	-0.02	0.16	0.55	0.07	0.51	0.26	0.74	0.00
Germany	0.18	0.43	0.10	0.69	0.38	0.81	0.16	0.19	0.55	0.07	0.50	0.26	0.74	0.00	0.20	0.59	0.07	0.60	0.24	0.75	-0.01
Italy	0.19	0.32	0.14	0.76	0.55	0.78	0.29	0.18	0.41	0.10	0.69	0.38	0.72	0.10	0.20	0.53	0.09	0.55	0.28	0.72	0.00
Japan	0.19	0.36	0.14	0.71	0.48	0.76	0.20	0.17	0.46	0.09	0.66	0.34	0.74	0.07	0.18	0.53	0.07	0.57	0.28	0.74	0.02
Panel C. Nonparametric specification																					
US	0.07	0.41	0.02	0.79	0.5	0.97	0.53	0.11	0.49	0.05	0.78	0.37	0.87	0.23	0.13	0.54	0.06	0.68	0.31	0.84	0.13
UK	0.09	0.39	0.06	0.78	0.51	0.92	0.40	0.10	0.37	0.07	0.77	0.48	0.82	0.26	0.11	0.43	0.06	0.71	0.44	0.82	0.24
Canada	0.08	0.45	0.04	0.76	0.49	0.94	0.42	0.08	0.42	0.05	0.78	0.44	0.86	0.27	0.10	0.44	0.07	0.74	0.40	0.82	0.18
France	0.17	0.75	0.04	0.60	0.20	0.94	0.17	0.16	0.64	0.06	0.68	0.23	0.86	0.09	0.18	0.75	0.06	0.60	0.15	0.86	0.00
Germany	0.19	0.52	0.09	0.68	0.37	0.88	0.25	0.19	0.51	0.09	0.68	0.31	0.80	0.10	0.22	0.57	0.10	0.60	0.27	0.77	0.03
Italy	0.18	0.50	0.08	0.71	0.41	0.89	0.32	0.17	0.46	0.08	0.72	0.38	0.81	0.18	0.20	0.50	0.11	0.61	0.33	0.76	0.08
Japan	0.18	0.61	0.05	0.66	0.32	0.92	0.27	0.17	0.51	0.07	0.73	0.32	0.83	0.15	0.19	0.53	0.08	0.69	0.30	0.82	0.11

Notes. For the Markov switching, linear and nonparametric approaches, and each forecasting horizon $h = 1, 2, 3$, the table shows the total Brier Score (BS), the Brier Score conditional to recessions (BSR) and to expansions (BSE), the area under the ROC curve $AUROC$, the average True Positive Rate (TPR), the average True Negative Rate (TNR) and the average Cohen's kappa coefficient (κ).

Table 4. Empirical performance in G7 countries: 1955.2-2022.3

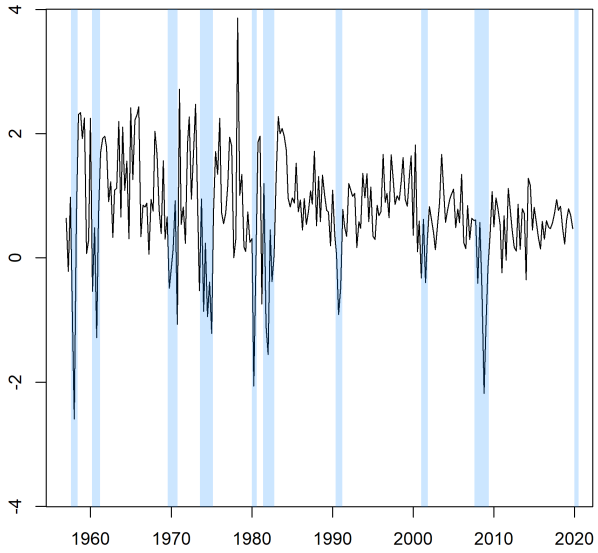
	h=1							h=2							h=3						
	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ	BS	BSR	BSE	$AUROC$	TPR	TNR	κ
Panel A. Markov-switching model																					
US	0.13	0.98	0.01	0.50	0.01	0.99	0.00	0.13	0.98	0.01	0.50	0.01	0.99	0.00	0.13	0.98	0.01	0.50	0.01	0.99	0.00
UK	0.12	0.99	0.01	0.50	0.01	0.99	-0.01	0.12	0.97	0.01	0.50	0.01	0.98	-0.01	0.12	0.96	0.01	0.50	0.02	0.98	-0.01
Canada	0.11	0.98	0.01	0.50	0.01	0.99	0.00	0.10	0.98	0.01	0.50	0.01	0.99	0.00	0.10	0.98	0.01	0.50	0.01	0.99	0.00
France	0.21	0.97	0.01	0.50	0.02	0.99	0.01	0.21	0.98	0.01	0.49	0.01	0.98	-0.01	0.21	0.97	0.01	0.49	0.01	0.98	-0.01
Germany	0.34	0.13	0.40	0.68	0.64	0.37	0.00	0.35	0.13	0.40	0.60	0.64	0.37	0.00	0.35	0.13	0.40	0.55	0.64	0.36	0.00
Italy	0.24	0.97	0.01	0.50	0.02	0.99	0.01	0.24	0.97	0.01	0.49	0.01	0.98	-0.01	0.24	0.96	0.01	0.49	0.02	0.97	-0.01
Japan	0.23	0.97	0.01	0.51	0.01	0.99	0.00	0.23	0.97	0.01	0.50	0.01	0.99	0.00	0.23	0.97	0.01	0.50	0.01	0.99	0.00
Panel B. Linear model																					
US	0.13	0.31	0.11	0.81	0.49	0.79	0.17	0.13	0.49	0.07	0.75	0.30	0.75	0.03	0.13	0.55	0.07	0.56	0.26	0.75	0.01
UK	0.14	0.55	0.08	0.59	0.27	0.81	0.03	0.13	0.70	0.06	0.69	0.17	0.78	-0.03	0.12	0.53	0.07	0.58	0.27	0.74	0.01
Canada	0.14	0.41	0.10	0.69	0.39	0.79	0.07	0.12	0.57	0.06	0.52	0.25	0.75	0.00	0.12	0.56	0.07	0.48	0.25	0.75	0.00
France	0.17	0.52	0.09	0.64	0.30	0.81	0.07	0.17	0.69	0.06	0.65	0.17	0.79	-0.04	0.16	0.53	0.07	0.54	0.27	0.74	0.00
Germany	0.19	0.54	0.07	0.66	0.29	0.84	0.11	0.20	0.62	0.07	0.62	0.21	0.76	-0.03	0.18	0.54	0.07	0.56	0.27	0.75	0.01
Italy	0.18	0.47	0.09	0.67	0.34	0.82	0.13	0.19	0.60	0.06	0.56	0.23	0.76	-0.01	0.19	0.55	0.07	0.53	0.25	0.75	0.00
Japan	0.18	0.43	0.11	0.70	0.40	0.79	0.16	0.17	0.54	0.06	0.64	0.27	0.77	0.03	0.18	0.55	0.06	0.59	0.26	0.76	0.01
Panel C. Nonparametric specification																					
US	0.08	0.42	0.03	0.78	0.50	0.96	0.49	0.12	0.49	0.06	0.77	0.36	0.86	0.20	0.14	0.56	0.07	0.67	0.30	0.83	0.11
UK	0.10	0.39	0.06	0.78	0.51	0.92	0.39	0.11	0.37	0.08	0.76	0.46	0.80	0.22	0.11	0.44	0.07	0.77	0.42	0.81	0.21
Canada	0.09	0.48	0.05	0.74	0.47	0.94	0.39	0.10	0.44	0.06	0.76	0.42	0.84	0.22	0.11	0.46	0.07	0.72	0.38	0.82	0.16
France	0.17	0.70	0.05	0.62	0.24	0.94	0.05	0.17	0.63	0.07	0.68	0.24	0.85	0.08	0.19	0.72	0.07	0.60	0.17	0.84	-0.01
Germany	0.20	0.52	0.09	0.68	0.37	0.88	0.25	0.20	0.50	0.10	0.68	0.32	0.79	0.09	0.22	0.57	0.11	0.60	0.26	0.77	0.02
Italy	0.18	0.50	0.08	0.71	0.41	0.89	0.32	0.18	0.44	0.09	0.72	0.39	0.80	0.18	0.21	0.51	0.11	0.61	0.32	0.76	0.07
Japan	0.19	0.62	0.06	0.65	0.31	0.91	0.25	0.17	0.49	0.07	0.73	0.34	0.82	0.15	0.19	0.55	0.08	0.69	0.29	0.81	0.09

Notes. For the Markov switching, linear and nonparametric approaches, and each forecasting horizon $h = 1, 2, 3$, the table shows the total Brier Score (BS), the Brier Score conditional to recessions (BSR) and to expansions (BSE), the area under the ROC curve $AUROC$, the average True Positive Rate (TPR), the average True Negative Rate (TNR) and the average Cohen's kappa coefficient (κ).

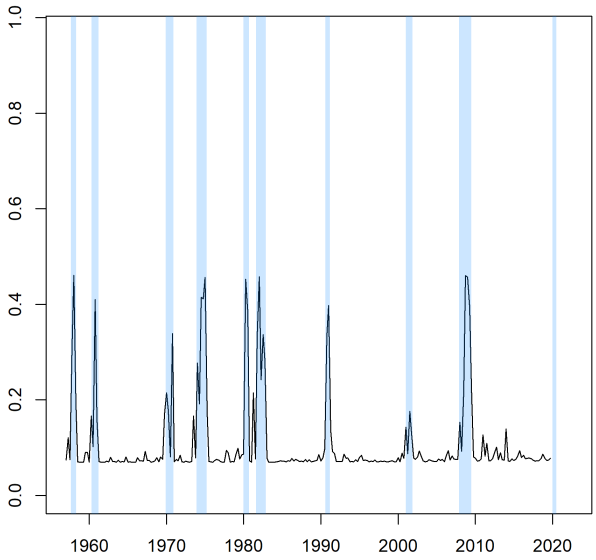
2.11 — Figures

Figure 2.1. Business cycle inferences: 1955.2-2019.4

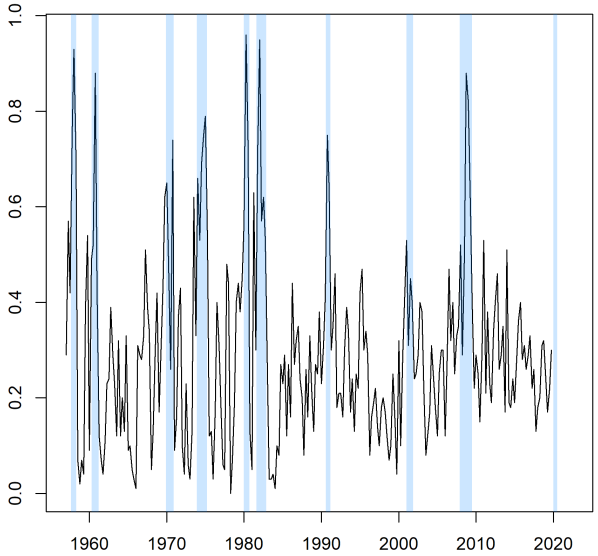
(a) Panel A. US GDP growth rates



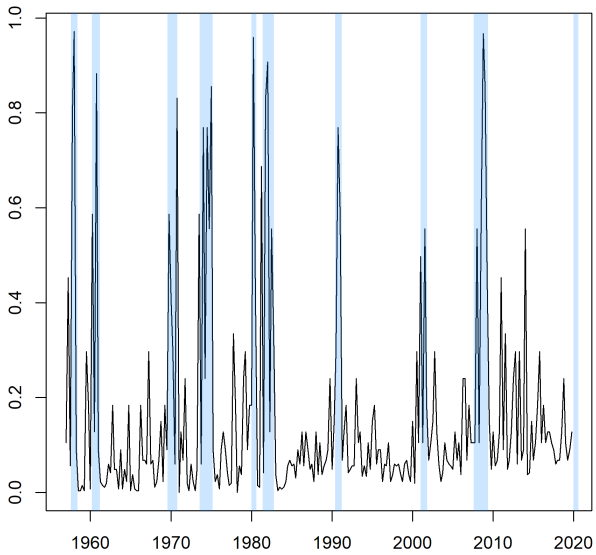
(b) Panel B. Probability from Markov-switching



(c) Panel C. Probability from AR(2)



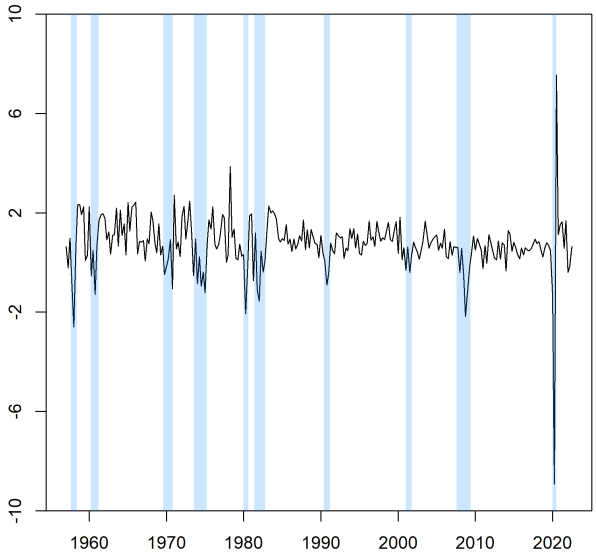
(d) Panel D. Probability from nonparametric



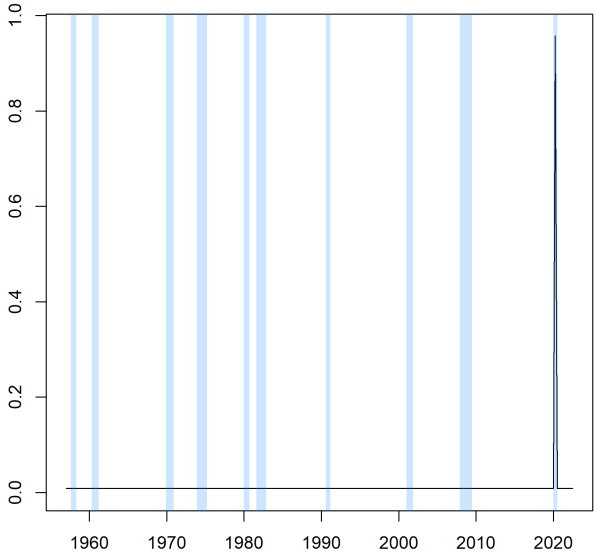
Notes. Panel A displays the US quarterly GDP growth rate for the period 1955.2-2019.4. Panels B, C, and D plot the 2-quarter ahead predictions of the probability of recession from AR(2), Markov-switching and nonparametric models, respectively. The shaded areas represent the NBER-referenced recessions.

Figure 2.2. Business cycle inferences: 1955.2-2022.3

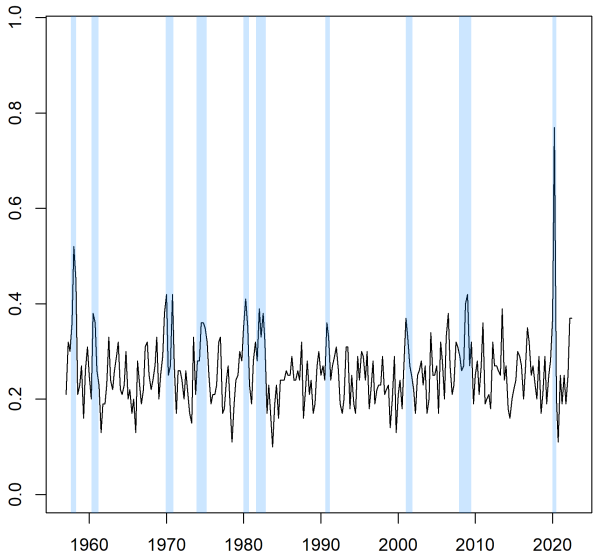
(a) Panel A. US GDP growth rates



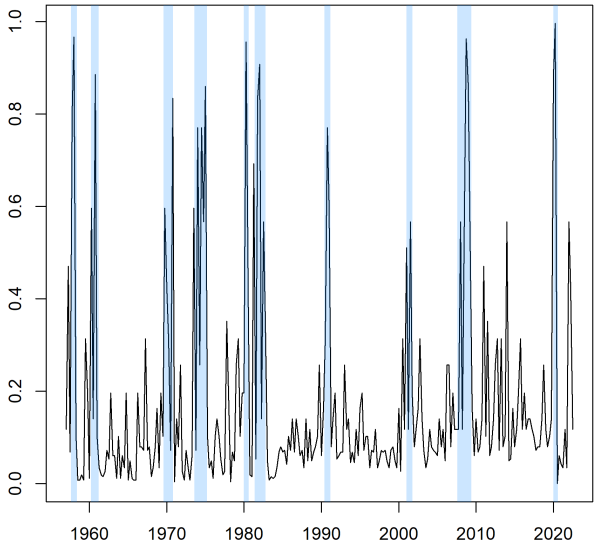
(b) Panel B. Probability from Markov-switching



(c) Panel C. Probability from AR(2)



(d) Panel D. Probability from nonparametric



Notes. Panel A displays the US quarterly GDP growth rate for the period 1955.2-2022.3. Panels B, C, and D plot the 2-quarter ahead predictions of the probability of recession from AR(2), Markov-switching and nonparametric models, respectively. The shaded areas represent the NBER-referenced recessions.

Chapter 3

What economic indicators point to recessions in Spain?

3.1 — Introduction

Economic growth in Spain, as in the rest of the industrialised countries, does not imply a continuous increase in the level of economic activity. On the contrary, as the National Bureau of Economic Research (NBER) points out, economic growth is characterised by a sequence of expansions and recessions. According to this institution's view, a recession, which starts at a peak and ends in a trough, implies a significant fall in the level of economic activity common to all economic sectors, lasting longer than a few months and typically seen in indicators such as GDP, income, employment, industrial production and sales. The NBER defines expansions symmetrically.

The Great Recession, which led to a general decline in the level of economic activity during the first decade of the 21st century, demonstrated the devastating power that recessions can have particularly in Spain. Stock market falls, loss of confidence among economic agents, problems in bank balance sheets, the uncontrolled rise in unemployment and the collapse in house prices are just some of the effects that this intense and prolonged recession had on the Spanish economy.[1]

For this reason, in recent years there have been numerous studies analyzing the characteristics of the economic cycle in Spain. On the one hand, various economic indicators have been proposed as a weighted average of individual indicators, as in Camacho and Pérez Quirós (2011), Camacho and Doménech (2012), Cuevas and Quilis (2012) and Cuevas, Pérez Quirós, and Quilis, (2017). On the other hand, the synchronization of the Spanish economy at the international level has been recently analyzed, as a country in Camacho, Caro, López-Buenache (2020) and by regions in Gadea-Rivas, Gomez-Loscos, and Bandrés (2017) and in Gadea-Rivas, Gomez-Loscos, and Leiva-León (2019). Finally, the synchronization of the Autonomous

[1] Camacho, Gadea-Rivas and Pérez Quirós (2019) rank the Spanish economy among those most negatively affected by the Great Recession.

Communities with the national cycle has been analyzed, as in Cancelo (2004), Gadea-Rivas, Gomez-Loscos and Montañés (2012), and in Camacho, Pacce, and Ulloa (2018).

Despite the progress made in understanding the characteristics of the Spanish economic cycle and its synchronization at an international and regional level, we believe that an exhaustive analysis of which are the best indicators to anticipate recessions in the Spanish economy is still to be carried out. Although the Spanish Economic Association decided to create a Spanish Economic Cycle Dating Committee in 2012 in charge of providing a historical dating of the reference cycle in Spain, the determination of the turning points in the phases of the cycle is done a posteriori so the dating of peaks and valleys is done with a considerable delay. Knowing the indicators that anticipate these phase changes in the Spanish economic cycle will be very useful for economic agents to make optimal decisions.

The framework we have chosen to perform in this chaptere this analysis is classification trees, which are within supervised learning algorithms. There are several reasons that justify this choice. The first reason is because they are a very simple and flexible nonparametric tool for classification of observations, allowing the use of categorical data and without much impact from outlier data. The second reason is that they allow to operate with very large databases since they work with very efficient algorithms from the computational point of view. These techniques are very stable even in cases where the number of explanatory variables exceeds the sample size. The third reason is because they are very easy to interpret as the results are usually accompanied by very intuitive graphical representations. The fourth reason is that they provide an ideal framework for examining the relative importance of indicators in the classification of recessions. In addition, classification trees allow us to examine the evolution of the relative importance over time of the indicators and thus help to discriminate which indicators best predict each of the recessions.

Recently, some papers have examined the ability of boosting-based classification trees for the prediction of economic recessions, such as Ng (2014) for the US, and Döpke, Fritsche, and Pierdzioch (2017) for Germany. Complementarily, Ward (2017) uses classification trees that are combined using random forest to identify international financial crises. Piger (2020) performs a comparison of both (and other) procedures to predict US recessions and concludes that the boosting technique produces better results. For this reason, we will use classification trees that are estimated using the Stochastic Gradient Boosting algorithm described in Friedman (2002).

The main results obtained in the empirical application are the following. First, the ability of the classification trees to anticipate the recessions of the Spanish economy 3- and 6-month ahead is very high, both in the in-sample exercise and in the recursive prediction. Second, in general, the ability of the leading indicators of the trend GDP and car sales, as well as of the registered unemployment and consumer prices series, is remarkable. Third, the predictive ability increases when we add the information contained in the confidence indicators, stock market, and interest rates. Fourth, in predicting the Great Recession, financial indicators

and leading indicators of the evolution of construction took on higher importance. Fifth, the current recession caused by Covid-19 is a production and employment crisis, where financial or construction variables are not so important this time around and where leading indicators of GDP, car sales or the registered unemployment and consumer prices series are once again the determining variables.

The structure of the chapter is as follows. In Section 2, we present the classification trees applied to the classification of a period in time in recession or expansion based on the information available at a previous point in time of a set of economic indicators. For readers uninitiated in boosting, we present the Adaboost algorithm and the (Stochastic) Gradient Boosting, both of which are based on iteratively improving the classification of simple trees by focusing on observations that have been difficult to classify up to the previous iteration. In Section 3, we present the results of the application of these techniques to the prediction of the recessions in the Spanish economy. Conclusions are then presented in the Section 4. This is followed by the references, appendix, and figures referred to in the chapter.

3.2 — Classification trees for predicting recessions

This section is devoted to presenting the notation and operation of classification trees, as well as to adapting this technique to the analysis To facilitate this purpose, a very simple economic example is included. Readers more familiar with these techniques can jump directly to the empirical application.

3.2.1. Introduction to classification trees

Applied to the prediction of the phases of the business cycle with h periods ahead, the classification problem consists of predicting at each moment of time t whether the economy will be in the recession or expansion group at $t + h$, based on the observation of K indicators $x_t = (x_{1t}, ..., x_{Kt})$, with $t = 1, ..., T$. The classification in the decision trees consists of partitioning the set of possible values of the indicators into regions that are classified into expansions or recessions for $t + h$ trying to minimize a loss function. As a result, the period $t + h$ will be classified in the phase of the cycle of the region to which x_t belongs.

The classification problem can be viewed as a hierarchical model of latent variables. Let us label the business cycle by an unobserved latent variable t s which takes values at time t in the set E, R , where E refers to expansions and R to recessions[2]. Let us call $S = s$ the set of realizations of the latent variable. Therefore, when at time t the latent variable takes a given value, we know the phase of the economic cycle in which the economy is: if $s_t = E$ the economy

[2]One could extend the approach to assume more than two possible states if needed.

is in expansion while if $s_t = R$ the economy is in recession.

To describe how classification trees work, we need to clarify some concepts. Let us call X the set containing all possible x_t values. In the process of classification, we need to partition the space X into J disjoint regions, $A_1, ..., A_J$, such that $X = \bigcup_{j=1}^{J} A_j$. The classification consists of assigning each vector x_t to a region A_j and proposing an estimate for the probability that the observations in that region are in recession. The determination of a region as an economic recession will be made when the probability of recession in the region exceeds a certain threshold, for instance, 0.5.

Suppose we know s_t for all $t = h+1, ..., T+h$. We can calculate the probability of a recession for any value of the indicators belonging to a region as the proportion of the observations in the region that come from a recession. If one defines I(•) as an indicator whose value is 1 if the condition inside the parentheses is satisfied and 0 otherwise, the estimate of the recession probability for a region A_j can be estimated as

$$p_j^R = \frac{1}{T} \sum_{x_t \in A_j} I(s_{t+h} = R), \tag{C.1}$$

where T_j is the number of observations in region j . The probability of expansion for any observation in that region will be $p_j^E = 1 - p_j^R$.

Once we have determined the probability of recession and expansion for each region, we can make an inference about the phase of the cycle in which the observations of the indicators belonging to each region will be found. A widespread option in the analysis of economic cycles is to present the inference that a model makes on the probability of recession at time $t+h$, once x_t has been observed. With classification trees, if we known about x_t , we can determine to which region it belongs and one will make inference about the probability of recession to which it results as $p(s_{t+h} = R|x_t \in A_j) = p_j^R$.

With this information, the user can determine the threshold ϑ from which to classify the time $t + h$ as a recession from the value x_t. To the inference we make about the phase of the cycle once the data are known, x_t, and the parameters used in the model, c, we will call them $\hat{s_{t+h}}(x_t, c)$ (we will omit from now own the dependence for clarity in the exposition). Therefore, the inference will be $\hat{s}_{t+h} = R$ if $p(s_{t+h} = R|x_t \in A_j) > \vartheta$ but it will be $\hat{s}_{t+h} = E$ if $p(s_{t+h} = R|x_t \in A_j) \leq \vartheta$. This way, we can infer the value of the realizations of the latent variable $\hat{S} = (\hat{s}_{h+1}, ..., \hat{s}_{T+h})$. In classification trees with two possible states, we usually take the threshold $\vartheta = 0.5$, i.e., we classify the region in the most probable phase of the cycle according to the observations that are most likely to occur to that region.

This method allows for easy analysis of results and prediction. First, the classification trees will allow us to understand the reasons why the moment of time $t + h$ has been classified as a recession or expansion by analyzing the characteristics of the region in which the vector of

observations of indicators x_t is located. Secondly, we can predict the probability of recession for any period τ, once we know the value of the indicators x_t, from the probability of recession of the region where x_τ belong to, $p(s_{\tau+h} = R|x_\tau \in A_j) = p_j^R$ even if $\tau > T$.

3.2.2. Creating the classification tree

Breiman et al. (1984) describe a procedure for partitioning decision trees known as Classification and Regression Trees (CART). This procedure is based on a classification rule or function defined on the space X such that, for each x_t , it indicates which region the observation belongs to, and in our case, makes inference about the assignment of each region to a phase of the economic cycle.

Suppose we know the set of realizations of the latent variable S. For the creation of the regions it is necessary to identify a loss function that allows us to compare some partitions with others. Since the inference on the state in which a region is classified is a function of the percentage of observations in the region that are in that state, the partitions are performed with the goal of finding the least impure regions as possible (for instance, if a region is classified as recession, it wiil be tried that it contains the largest percentage of observations in recessions as possible). To perform the partitions a measure of the degree of impurity of a tree and a rule that allows us to perform partitions are needed.

A measure that allows us to determine the degree of impurity of a region A_j is the Gini index (others are classification error or entropy),

$$G_j = \sum_{s=E,R} p_j^s(1 - p_j^s) = 2p_j^E p_j^R. \tag{C.2}$$

The index takes the maximum value of 0.5 when the distribution of cycle phases in the region is very impure, i.e. the region contains the same number of recessions as expansions. However, the index takes values closer to zero the more predominant one of the two states is. The degree of impurity of the tree will be measured as the weighted average of the degree of impurity of each of its regions.

$$G_= \sum_{j=1}^{J} \frac{T_j}{T} G_j, \tag{C.3}$$

where the weights reflect the weight of the region's observations over the total. This indicator is a measure of the total variance of the classification tree.

Since it is not feasible to consider all possible partitions of space X , we use the *recursive binary splitting* algorithm, which is based on performing a recursive partitioning. At the begin-

ning of the algorithm, all observations are in a single region, A , which coincides with the entire space X. The partitions we are going to consider are rectangular (although non orthogonals as in Paez et al. (2019) can be applied) and are going to be formed from thresholds $c = (c_1, ..., c_K)$ that we will compare with the values that can be taken by the indicators $(x_1, ..., x_K)$.

To perform the first partitioning, we will select an indicator x_k and a threshold c_k^1 that will give rise to two rectangular regions $A_1^1 = \{X|x_k < c_k^1\}$ and $A_2^1 = \{X|x_k \leq c_k^1\}$. For the first region, the indicator x_k will take values below the threshold c_k^1 . For the second region, the indicator will be greater than or equal to the threshold. For each indicator x_k and each threshold c_k^1 we will calculate the Gini index and choose the indicator and the threshold that results in the partition that minimizes the degree of impurity of the resulting tree.

For the second partition, we will need to divide one of the previous two regions into two, resulting in a segmentation of X into three regions. The two possibilities are either $A_1^2 = \{X|x_k < c_k^1, x_j < c_j^2\}$, $A_2^2 = \{X|x_k < c_k^1, x_j \geq c_k^2\}$ y $A_3^2 = \{X|x_k \geq c_k^1\}$, or $A_1^2 = \{X|x_k < c_k^1\}$, $A_2^2 = \{X|x_k \geq c_k^1, x_j < c_j^2\}$ and $A_3^2 = \{X|x_k \geq c_k^1, x_j \geq c_j^2\}$. Again, for each indicator x_j and each threshold c_k^2 we will calculate the Gini index in each of the two possibilities above. We will keep the partition such that the new indicator and the new threshold result in a partition with the lowest index value.

In this recursive procedure, the partition generated in one step will give rise to two new regions in the next step. This process is repeated iteratively until some criterion is reached that stops the algorithm. The idea is that trees with many partitions will have a low degree of impurity, but will have little ability to classify new observations well. Some of the most common criteria for limiting the number of partitions are that no region should contain less than a minimum number of observations, that the space X should have a maximum number of partitions or that the realization of a new partition does not imply much reduction in the reduction in the Gini index of the resulting tree. The most common option is starting from a tree that gives rise to many regions and "pruning" the tree in order to minimize the Gini index penalized by the number of regions using cross-valuation procedures to determine the degree of penalty.

Figure 3.1 allows us to understand the idea behind the regression trees by means of a simplified example, in which we have assumed $h = 0$. The objective of the example is to create a classification tree to determine whether in one period the Spanish economy is in recession or in expansion. To determine the state variable of the business cycle, $S = s_1, ..., s_T$, we have used the quarterly reference economic cycle determined by the Spanish Business Cycle Dating Committee (CFCEE) or the period between 1990.2 and 2019.2. Therefore, $s_t = Rec$ indicates that the Committee determined that in quarter t the Spanish economy was in recession, while $s_t = Exp$ would indicate that the economy was in expansion.

The economic indicators x_t that have been taken into account to prepare the classification

tree are the seasonally adjusted unemployment rate of the Labor Force Survey and the growth rate of the Industrial Production Index (IPI), which has been quarterly adjusted by taking the quarterly average. Using these indicators and the state variable, we have estimated a decision tree by imposing four partitions and we have plotted the result in Figure 3.1.

This figure shows the elements that make up the usual terminology of decision trees. The root node represents the entire space X and is sequentially partitioned into new regions up to the four final regions shown at the bottom. The various branches of the tree arise from the root node, two new ones at each of the decision nodes. Finally, the terminal nodes, or leaves of the tree represent the four regions into which the space has been divided. For ease of interpretation, the nodes report the percentage of recessions and the percentage of observations that include the regions they determine.

The top panel of Figure 3.1 graphically represents the tree structure with its nodes, branches and leaves. At the beginning, the tree assigns the observations to expansion because only 21% of this sample were recessions. The first partition assigns 11% of the observations that have an IPI growth rate below 1.73% to the right branch and classifies them as recessions since 85% of these observations are recessions. This region is not re-partitioned. Eighty-nine percent of the observations whose IPI growth rate is above 1.73% are assigned to the left branch of the tree and this is the branch that undergoes with new partitions. The probability of recession for this region is only 12% since this is the percentage of recessions that contains.

In the lower panel of Figure 3.1, the four partitions of the X space that the classification tree has given rise to are represented. Each of the points depicted in the plane refers to each of the T=117 observations of $x_t = (IPI_t, Unemp_t)$. The 93 observations that have been classified by the Cycle Dating Committee (CFCEE by its acronym in spanish) as expansion are shown in red, while the 24 observations classified as recessions by the Cycle Dating Committee are shown in blue. The first region, located on the left, is characterized by the value taken by industrial production, $A_1^3 = X|IPI < -1.73$, and any observation that belonging to it will be classified by the tree as a recession.

The second region, located on the lower right, is characterized by a combination of not very low industry growth and a low unemployment rate, $A_2^3 = X|IPI \geq -1.73, Unemp. < 17.31$, and the tree infers an expansion for it. The two regions in the upper right are characterized by high unemployment (equal to or greater than 17.31). The leftmost region also has very low growth in industrial production and is classified as a recession,
$A_3^3 = X| - 1.73 < IPI < -0.29, Unemp. \geq 17.31$, while the one further to the right offers a not so low industrial production growth, $A_4^3 = X|IPI \geq -0.29, Unemp. \geq 17.31$, and is classified as expansion. Therefore, in this example recessions in the Spanish economy are characterized by a sharp fall in industrial production or a somewhat more moderate fall in industrial production combined with high unemployment.

To measure the degree of tree purity we can use the Gini index. The region A_2^3 is very pure in the sense that it only contains observations that have been classified by the CFCEE as expansions, so the Gini index is minimal and is $G_4 = 0$. The purity of regions A_1^3 and A_4^3 is also very high: for A_1^3 10 of the 12 observations came from recessions ($G_1 = 0.28$), while for A_4^3, 34 of the 39 observations came from expansions ($G_4 = 0.22$). The most impure region is A_3^3, as only 8 out of 13 observations come from recessions ($G_3 = 0.47$). The total Gini index, as a weighted average of the above, is $G = 0.16$.

Finally, it is interesting to measure not only the degree of purity of the tree but also its ability to form regions that do not contain misclassified observations. A simple way to measure the classification ability of the tree is to compare the value of the realizations of the latent variable $S = s_1, ..., s_T$ with their inference using the classification tree $\hat{S} = \hat{s}_1, ..., \hat{s}_T$ such that $\hat{s}_t = \hat{s}_t(x_t | x_t \in A_m)$ is the classification that is made for the values of the indicators that belong to the region A_m (this is, s_t is external and does not depend on indicators, while $\hat{s}_t$ is inferred based on the region where the indicators x_t where classified).

Thus, region A_1^3 contains two observations of expansions. For this region recession was inferred $\{\hat{s}_t = Rec | x_t \in A_1^3\}$ because most of the moments in time t (with values of the indicators x_t) for that region were determined by the Committee as economic recessions. In region A_2^3 expansions are inferred, $\{\hat{s}_t = Exp | x_t \in A_2^3\}$, and does not contain observations of recessions. In region A_3^3 recession is the classification, $\{\hat{s}_t = Rec | x_t \in A_3^3\}$, although five observations of expansions appear. Finally, the region A_3^3 contains five recession observations when it is a region of expansions, $\{\hat{s}_t = Exp | x_t \in A_4^3\}$. Therefore, the percentage of observations misclassified by the tree, out of the total number of observations, is 10%.

The graph is very intuitive because the resulting regions have been shaded as a function of the recession probability of the observations in each region. The areas that appear darker indicate a higher probability of recession. Thus, in a very visual way we can visually determine the probability of recession for any combination of IPI and unemployment rate. Finally, Breiman et al. (1983) propose a very useful measure to select the relative importance of indicators for ranking relative importance of indicators to classify observations between expansions and recessions. The relative importance of the indicator x_k in the final ranking is determined by the number of times that indicator has been used to make the partitions, weighted by the reduction in the Gini index provided by the partitions in which it participates.

Let's suppose that in a tree a total of P sub-partitions are made at each non-terminal node and that the indicator used to partition the tree is v_p , with $p = 1, ..., P$. If we name ΔG_p the reduction in the Gini index that occurs in that partition, the relative importance of the indicator x_k in the classification is

$$I_k = \sum_p^P \Delta G_p I(v_p = x_k), \tag{C.4}$$

where $k = 1, ..., K$. For ease of interpretation, the indicator is usually normalized so that the relative importance sums to 100. Figure 3.2 shows that the relative importance of industrial production for classification purposes is much higher than that of unemployment, since their relative importance is 83 and 17, respectively.

3.2.3. Results evaluation

To get to know the true classification ability of the tree we must examine to what extent it is able to classify observations that have not been used to generate the tree. As trees fall under what we know as supervised learning methods based on past experience, it is useful to divide the sample period into two sub-periods. The first subperiod is the training subperiod, for which we assume that we know the classification $S_{T_1} = s_{h+1}, ..., s_{T_1+h}$ and economic indicators data x_t for $t = 1, ..., T_1$, with $T_1 < T$ (we will choose it at the beginning of the sample without loss of generality).

The second subperiod is the evaluation one, in which the classification ability of the model is examined for observations that have not been used to generate the tree. In this case, the classification $S_{T_2} = s_{T_1+h+1}, ..., s_T$ will be assumed to be unknown and we will use the tree with observations up to T_1 to make a classification of the observations of the evaluation period $\hat{S}_{T_2} = \hat{s}_{T_1+h+1}, ..., \hat{s}_T$ from the indicators x_t for $t = T_{1+1}, ..., T$. The comparison between S_{T_2} and $\hat{S}_{T_2}$ will inform us of the true capacity of the tree to classify the observations.

3.2.4. Classification trees and boosting

Two problems have been identified in classification tree applications. The first problem is that trees often have limited classification ability in the evaluation period. The second is that they tend to produce results that are not very robust, in the sense that small changes in the data can produce large changes in the estimated classification tree. To alleviate these problems, one of the most widespread techniques is boosting, which began with a set of binary classification techniques proposed by Freund and Schapire (1996), based on the earlier work of Schapire (1990) and Freund (1995), and which are known as the AdaBoost (Adaptive Boosting) algorithm.[3]

The idea behind the application of boosting to classification trees is to sequentially create trees in which each new tree is a modification of the previous one, so that in each new iteration the algorithm learns from the mistakes made up to the previous iteration. The algorithm uses

[3]Other techniques used to make classification trees robust are bagging and random forest. Unlike boosting, they are based on generating copies of the original data using bootstrap. In each copy a classification tree is estimated and the results are combined to produce a final classification. Thus, deep trees are generated with little impurity, but with little ability to correctly classify new observations. However, by aggregating the results of many generated trees, the ability to classify new observations is improved without greatly increasing the impurity.

trees with few partitions, so they are very impure, although with a high capacity to classify new observations. However, as the trees are adjusted sequentially, the degree of impurity is reduced without losing classification ability. This is achieved because in each iteration the algorithm focuses on correctly predicting the observations that previous iterations have not been able to classify.

Discrete Adaboost

To apply Adaboost on classification trees, Ng (2014) describes the following version of the algorithm known as *discrete Adaboost*, which is one of the most common Adaboost algorithms in classification trees. For convenience, these techniques encode the state variable s_{t+h} in the set $\{-1, 1\}$,where $s_{t+h} = -1$ refers to the expansions and $s_{t+h} = 1$ refers to the expansions with $t = 1, ..., T - h$. The algorithm can be summarized as follows:

1. We start from equal weights for all indicator values $w_t^1 = \frac{1}{T-h}$. Let us define the classification up to iteration j as S_{t+h}^j, with $t = 1, ..., T - h$, to which we assign a neutral initial value of $S_{t+h}^1 = 0$ for all $t = 1, ..., T - h$.

2. For a set of iterations $m = 1, ..., M$, the classifications performed up to the previous iteration S_{t+h}^{m-1} are updated with those of the current iteration s_{t+h}^m adjusted by a correction factor α^m

$$\hat{S}_{t+h}^m = \hat{S}_{t+h}^{m-1} + \alpha^m \hat{s}_{t+h}^m. \tag{C.5}$$

   The correction parameter α^m is defined in the algorithm, by iteration of the following steps:

   (a) In iteration m we estimate a decision tree with a few partitions that will result in a classification $\hat{s}_{t+h}^m$, with $t = 1, ..., T - h$, in such a way to minimize the error that is made with this tree. In this case, the error is defined as the weighted sum of the number of times that observations are misclassified,

$$e^m = \sum_{t=1}^{T-h} w_{t+h}^m I(s_{t+h} \neq \hat{s}_{t+h}^m), \tag{C.6}$$

   The algorithm only updates $\hat{S}_{t+h}^m$ if $e^m < 0.5$, since $e^m = 0.5$ would be a random classification.

   (b) In such a case, we calculate the weight that we will give to the classification we obtain with this tree in the final classification by means of a function that decreases with the error made

$$\alpha^m = \frac{1}{2} ln(\frac{1 - e^m}{e^m}). \tag{C.7}$$

If the error tends to the maximum of 0.5, α^m tends to zero and that tree that generates the classifications $\hat{s}^m_{t+h}$ will have almost no influence on the final classification. The weighting of the tree estimated in step m will be higher the lower the classification error of that tree.

(c) We update the weight we will give to the observations in the next iteration as follows

$$w^{m+1}_{t+h} = \frac{w^m_{t+h}}{2\sqrt{e^m(1-e^m)}} exp(-\alpha^m s_{t+h}\hat{s}_{t+h}), \tag{C.8}$$

where it is satisfied that $\sum_{t=1}^{T-h} w^{m+1}_{t+h}$. When the tree classification error approaches the maximum error of 0.5, then $2\sqrt{e^m(1-e^m)}$ tends to 1 and α^m to zero. This implies that $w^{m+1}_{t+h} \simeq w^m_{t+h}$ and the weights of the observations to be used in the next iteration are almost not updated with that tree.

The algorithm will give more importance in iteration $m+1$ to the observations that did not classified well in iteration m. If the observation $t+h$ is well classified, the sign of $s_{t+h}\hat{s}_{t+h}$ will be positive and the weight will be multiplied by a number less than 1, which will be smaller the smaller the error of the tree is. Otherwise, the weight of that misclassified observation will be multiplied by a number greater than 1, which will be larger the smaller the error of the tree.

3. The final classification will be based on the average of the classifications of the trees constructed in the process, weighted in a decreasing way by the error they have made

$$\hat{s}_{t+h} = sign(\hat{S}^M_{t+h} = \sum_{m=1}^{M} \alpha_m \hat{s}^m_{t+h}), \tag{C.9}$$

where $sign(z) = 1$ if $z > 0$ and $sign(z) = -1$ if $z < 0$. Therefore, we can interpret that the classification of an observation in t will be determined by the decision of the qualified majority of classifications in the iterations. If the majority of times we have classified observation $t+h$ as an expansion ($\hat{s}^m_{t+h} = -1$) and we got it right (α_m large) the sign that will predominate for the weighted mean will be negative and then it will be finally classified as an expansion, since $\hat{s}_{t+h} = -1$.

In applications of this algorithm, it has been observed that the classification ability remains more or less constant for relatively small values of M when the trees used to generate $\hat{s}^m_{t+h}$ in step a) above are the result of minimizing the classification error of trees from only two regions. For more complex trees that generate more regions, the algorithm tends to converge much earlier, albeit at the cost of increased computational complexity. In either case, we can determine the classification error obtained at each iteration of the *discrete Adaboost* algorithm up to a value of M^* large and obtain the optimal number M of trees.

Friedman, Hastie, and Tibshirani (2000), provide statistical arguments to understand

why Adaboost algorithms give such good empirical results. In this influential contribution, they propose that it is easier to understand the process of classifiying observations between recessions and expansions as the search for the classifier S_t which minimizes the expectation of an exponential loss function

$$E(FP(S_{t+h})) = E[exp(-s_{t+h}S_{t+h})]. \tag{C.10}$$

When S_{t+h} classifies right, the signs of s_{t+h} and S_{t+h} coincide and the loss function is small. However, when S_{t+h} misclassifies, the loss function is large. Deriving the above expression and setting it equal to zero, it can be shown that the value of the classifier S_{t+h} that minimizes the above expectation is

$$S^*_{t+h} = \frac{1}{2} ln[\frac{p(s_t = 1|x_t)}{p(s_t = -1|x_t)}], \tag{C.11}$$

which coincides with half the logarithm of the odds ratio or ratio between the probability of a recession taking place and the probability of an expansion ocurring. Similar to logistic models, the conditional logistic models, the conditional probability that a recession occurs is

$$p(s_{t+h} = 1|x_t) = \frac{exp(2S_{t+h})}{1 + exp(2S_{t+h})} = \frac{exp(S_{t+h})}{exp(-S_{t+h}) + exp(S_{t+h})}. \tag{C.12}$$

In logistic models, the logarithm of the odds ratio is estimated as a weighted sum of the observations by the regression coefficients. Friedman et al. (2000) propose to estimate S^*_{t+h} using a nonparametric approach based on additive regression models

$$\hat{S}^M_{t+h} = \sum_{m=1}^{M} \alpha_m \hat{s}^m_{t+h}(c_m), \tag{C.13}$$

where it has been made explicit that the classifier, $\hat{s}^m_{t+h}$, depends on the thresholds, c^m, which need to be estimated for the partitions. The optimization problem involves finding numerically the value $\hat{S}^M_{t+h}$ that minimizes the exponential cost function (Eq.C.10), which implies finding the sequence of optimum weights and thresholds ${\alpha_m, c_m}_{m=1}^M$. From a given value of $\hat{S}^{m-1}_{t+h}$, sequential optimization algorithms applied in additive regression models are based on sequentially finding the pair α_m, c_m which minimizes

$$\alpha_m, c_m = argMin \sum_{m=1}^{M} exp(-s_{t+h}(\hat{S}^{m-1}_{t+h} + \alpha_m \hat{s}^m_{t+h}(c_m))). \tag{C.14}$$

Friedman et al. (2000) show that the optimum classifier, $\hat{S}^m_{t+h}(c_m)$ is the one that minimizes the weighted error shown in expression C.6, that the weights that minimize the exponential loss function, α_m, coincide with those in C.7, and also that the optimal update of the error weights is provided by C.8.

(Stochastic) Gradient Boosting

Friedman (2001) proposes an algorithm known as *Gradient Boosting* which generalizes the Adaboost algorithms proposed by Friedman et al. (2000). As prevously mentioned, AdaBoost focuses on applying simple trees to observations weighted by the error up to the previous iteration. In contrast, Gradient Boosting uses the algorithm *Functional Gradient Descent* algorithm to iteratively find the minimum of a function so that, at each iteration, the direction and size of update aimed at optimize the reduction of the function value is searched. Specifically, at each iteration the function updated in proportion to the negative of the gradient of the loss function that marks the direction of the update. For this reason, in practice, Gradient Boosting applies simple trees to try to approximate the negative of the gradient of the loss function evaluated in the classification of the previous iteration.

If we use an exponential loss function as in Eq.C.10, the objective is to find S_t that minimizes the loss function using a nonparametric approximation based on additive regression models $\hat{S}^M_{t+h}$ as in Eq.C.13. Following Schapire and Freund (2012), the algorithm consists of the iterative application of two steps. Starting from a classification $\hat{S}^{m-1}_{t+h}$, the first step in each iteration m we must calculate the direction of the update that is marked by the negative of the derivative of the exponential loss function evaluated at $\hat{S}^{m-1}_{t+h}$

$$g_{t+h} = \frac{\partial FP(S_{t+h})}{\partial S_{t+h}}\Big|_{S_{t+h}=\hat{S}^{m-1}_{t+h}} = s_{t+h} exp(-s_{t+h}\hat{S}^{m-1}_{t+h}). \tag{C.15}$$

The values of this expression are known as pseudo-residuals and will be larger for misclassified observations. Omitting the normalization, we can find the tree $\hat{s}^m_{t+h}$ that best fits the pseudo-residuals as the one that maximizes the function

$$PE = \frac{1}{T-h}\sum_{t=1}^{T-h} s_{t+h}\hat{s}^m_{t+h}(c_m) exp(-s_{t+h}\hat{S}^{m-1}_{t+h}). \tag{C.16}$$

It must be noted that maximizing this function is equivalent to minimize the error (6) in *Adaboost*. Moreover a quadratic loss function could be used to choose the tree minimizing the distance with the gradient, although the impact will not be significant in $\hat{S}^M_{t+h}$.

Gradient Boosting also puts emphasis on classifying well in iteration m the observations that have been misclassified in the previous iteration. An observation misclassified at iteration $m-1$ will imply a high value for $exp(-s_{t+h}\hat{S}^{m-1}_{t+h})$, and if it were to be misclassified again the sign of $s_{t+h}\hat{s}^m_{t+h}(c_m)$ would be negative so it would imply a burden to maximizing Eq.C.15.

Once the direction of the improvement has been estimated with $\hat{s}^m_{t+h}$, the second step consists of determining the size of the adjustment. Specifically, α_m is chosen as the value which minimizes the exponential loss function and we know from Friedman et al. (2000) that

it coincides with the expression C.7. Alternatively, Friedman (2001) proposes that, using the criterion of minimization of the loss function, a distinct weight α_{mj} could be found for each region j of the estimated tree in the iteration m. Another option is to assume that the weight of the trees at all iterations is a constant, $\alpha_m = \alpha$ for all m.

Lastly, Friedman (2001) incorporates a tuning parameter that controls the size of the jumps in the algorithm,

$$\hat{S}^m_{t+h} = \hat{S}^{m-1}_{t+h} + \eta\alpha_m \hat{s}^m_{t+h}(c_m), \tag{C.17}$$

where $0 < \eta \leq 1$. The smaller the value of the fit parameter, the better the classification in the training period, but the greater the number of iterations needed to reach the optimum. On the other hand, very large values may underestimate the number of trees needed. The empirical results of Friedman (2001) suggest to use values $\eta \leq 0.1$.

Finally, Friedman (2002) will incorporate an additional element in the *Stochastic Gradient Boosting* algorithm: the random sampling of a percentage ν of observations that are part of the training period. Specifically, at each iteration of the algorithm, the new decision tree is adjusted using only a fraction θ of data from the training period, extracted randomly from the training period. This procedure improves the classification capacity and speeds up the computation. In practice, to reduce computation time, θ will be smaller the greater the number of available indicators is.

3.2.5. Relative importance and interaction effect

In order that the classification trees can be used to automatically select the most influential indicators in anticipating the economic cycle, Friedman (2001) proposes a measure of relative importance in the boosting algorithm. Let I^m_k be the relative importance of the indicator x_k in each m tree out of the M generated by boosting, and obtained using expression (4). The relative importance of the indicator k x in the classification obtained with the boosting will be the average of its relative importance in the M trees generated:

$$IB_K = \frac{1}{M}\sum_{p}^{P} I^m_k, \tag{C.18}$$

where $k = 1, ..., K$. Economic indicators that are never used for partitioning will be discarded as business cycle indicator variables.

Finally, it is important to have some tool to clarify the black box that might appear to be the result of the decision trees estimated with boosting. To examine the role played by the indicators in the construction of the tree, Friedman (2001) proposes the use of Partial

Dependence Graphs. These graphs show the value taken by the result of the classification tree $\hat{S}^M_{t+h}$ for different values of some economic indicators. One must recall that $\hat{S}^M_{t+h}$ is an estimate of the half logarithm of the odds ratio and that, therefore, a higher value on the ordinate axis is related to a higher probability of recession. Therefore, these graphs will help us to examine whether the relationship between the probability of recession and economic indicators is null, linear, or more complex.

Without loss of generality, let us suppose that we want to measure the effect of the first indicator x_1 on $\hat{S}^M_{t+h}(x_{1t}, x_{2t}, ..., x_{Kt})$. The first step is to build the decision tree using boosting for the original database. In the second step, we replace the value of x_1 by a succession of possible values of the indicator $x_{1i} = x_{11}, x_{12}, ..., x_{1N}$. In the third step, we estimate $\hat{S}^M_{t+h}(x_{1t}, x_{2t}, ..., x_{Kt})$ for all $t = 1, ..., T-h$ and for each $i = 1, ..., N$. Finally, we calculate the mean

$$\bar{\hat{S}}^M_{t+h}(x_{1i}) = \frac{1}{T-h} \sum_{t=1}^{T-h} \hat{S}^M_{t+h}(x_{1i}, x_{2t}, ..., x_{Kt}), \tag{C.19}$$

with $i = 1, ..., N$. The graph of $x_{1i}, \bar{\hat{S}}^M_{t+h}(x_{1i})$ for $i = 1, ..., N$ is the Partial Dependence Plot for x_1.

Using boostrap techniques, confidence intervals can be estimated for the Partial Dependence Graphs. For this purpose, a number of randomly constructed B subsamples that include a percentage μ of the original data can be randomly constructed. For each of these samples, the mean response $\bar{\hat{S}}^{Mb}_{t+h}(x_{1i})$, with $b = 1, ..., B$, is calculated. If one works at a confidence level of confidence level of $\lambda\%$, the partial dependence will be the median of the observations and appears in the graph next to the quantiles 0.5λ and $(100 - 0.5\lambda)\%$.

In a simple way, the Partial Dependence Graph can be generalized for any set of x_t indicators. Although graphs of complex combinations of indicators are difficult to interpret, the most frequent combination is the one that measures the interaction effect between two indicators. In order to measure the interaction effect between two indicators a and b , the sequence $x_{ai}, x_{bi}, \bar{\hat{S}}^M_{t+h}(x_{ai}, x_{bi})$ for $i = 1, ..., N$ is used to be represented.

3.3 — Empirical application

For the empirical analysis we have selected a sample of 270 monthly economic indicators. After adjusting the observations, the effective sample period is from 1971.01 to 2020.03. However, many of the Spanish economic indicators start to be constructed from more recent dates, which prevents us from performing the analysis with all the indicators for the entire sample period. For this reason, we have divided the analysis into four partial studies, using respectively indicators starting in 1971, 1978, in 1988 and in 2004. We will refer to these four studies as 71s, 78s, 88s and 04s, respectively.

Specifically, economic indicators were obtained from the OECD's Main Economic Indicators (MEI) database, while indicators related to interest rates were from the International Monetary Fund (IFS), and monetary aggregates data from Bloomberg. When necessary, the indicators have been conveniently transformed to work with their stationary versions. The number of indicators used to carry out these analyses is growing. In the 71s analysis, 36 indicators are used, including the OECD leading indicators. OECD leading indicators related to production, car sales, capacity utilization, employment, employment capacity utilization, employment, prices, exchange rates and trade.

Analysis 78s incorporates 39 new OECD indicators related to construction and interest rates, in addition to other production, employment and price series. In the analysis 88s, 58 new consumer confidence, stock market and interest rate indicators,as well as new production, prices, trade and employment indicators. The largest expansion of indicators appears in the analysis of 04s with 137 new indicators, mainly for employment, prices and interest rates. The importance of the analysis carried out with these 270 indicators is that we can make inferences about the probability of recession during the on the probability of recession during the Great Recession with a large battery of indicators, as well as for the recent Covid-19 crisis.

To establish the recession periods we have used the Spanish benchmark business cycle dating provided by the Spanish Business Cycle Dating Committee created by the Spanish Economic Association in 2012. In the sample used, the Committee classifies five recession periods that account for 13% of the observations in this sample.

The analysis of the predictive capacity of the indicators has been made at prediction horizons h of 3 and 6 months. The estimation is performed using the GBM package of R. We have used the "adaboost" loss function. The maximum number of trees allowed is 2000 and the optimum M has been chosen by five-fold cross validation. The minimum number of observations per node is 5. The minimum number of observations per node is 5, and 6 final regions are built in the trees to allow for the interaction effect between the indicators. Boosting adjustment parameter is = 0.005 Boosting is $\eta = 0.005$, while in Stochastic Gradient Boosting the fraction θ of data from the training period is 50%. To make inference in the partial dependence plots we have used a total of B = 0.005. partial dependence plots we have used a total of B = 1000 bootstrap replications.

3.3.1. In sample analysis

The relative importance of the economic indicators is shown in Figure 3.3. As an illustration, we will focus on the cases that include more sample period (71s) and more economic indicators (04s). For ease of interpretation, the importances have been scaled to sum up to 100 and only the 10 indicators with the highest relative importance are shown. Using the 71s analysis, the main indicators are those of the OECD leading indicators for the evolution of GDP, unemployment

and car sales. If we focus on the 04s analysis, which only includes the two recessions that are known as the Great Recession, production and the leading indicator for construction, price indexes, and financial indicators gain prominence. However, car sales remain important.

Using the relative importance criterion of the 70s analysis, Figure 3.4 shows the evolution (top plots) and partial dependence (bottom plots) of two of the most important economic indicators: the year-on-year growth rates of the leading indicator of car sales and registered unemployment. (the upper graphs show shaded areas marking the periods Cycle Dating Committee classifies as recessions along with the evolution of the indicators while in the lower graphs, along with the partial dependence, the 95% confidence intervals are included as shaded areas). Appendix A shows two other possible applications of relative importance to the study of business cycles, on the one hand, to analyze which country is the most relevant for predicting the probability of recession in the world cycle (with particular reference to the OECD), and on the other hand, to develop a measure of synchronization between two countries, with particular reference to the case between the United States and Spain.

In the partial dependence plots, the value of the economic indicators is shown on the x-axis while half of the logarithms of the odds ratio at a 6-month prediction horizon for those values and their respective confidence intervals are shown on the ordinate axis. As one would expect, the relationship between the value of the indicators and the probability of recession is negative in the case of the leading indicator of car sales, while the relationship is positive in the case of the unemployment indicator. In all cases, there is a clear non-linear effect of the indicators on the estimated probability of recession. In the case of the car sales indicator, the lower left graph shows a drop in the probability of recession(measured on the odds ratio scale) when the year-on-year rate of the indicator falls below 20%. The upper left graph shows that these negative rates of the indicator coincide with the Committee's recessions. In the case of unemployment, the steep rise in the probability of recession appears for rates close to 40% in the lower right graph. The upper right graph shows that these values are correlated with the Committee's recessions.

To illustrate the importance of the interaction effect between economic indicators, Figure 3.5 shows the effect on the logarithm of the odds ratio (right scale) at a 6-month forecast horizon of the joint values of the growth rates of the leading indicator of car sales (y-axis) and of registered unemployment (abscissa axis). The interpretation of the graph is very simple: the lighter the shaded area, the greater the probability of recession (measured on the odds ratio scale) of the combination of leading indicator and employment values is. For a given value of the change in unemployment, the color of the area becomes lighter when the leading indicator value of car sales falls, while, for a given value of the year-on-year change in the leading indicator of car sales, the area is lighter the higher the growth of unemployment is. However, the combination of low values of the year-on-year rate of the leading indicator of car sales and high values of the year-on-year change in the unemployment indicator are especially

alarming in terms of the probability of recession.

Figure 3.6 shows the inference on the estimate of the recession probability $p(s_{t+h} = 1|x_t)$ obtained by models 71s, 78s, 88s and 04s for prediction horizons h of 3 and 6 months. In addition, shaded areas are included to mark the recessions of the Spanish economy, according to the Spanish Business Cycle Dating Committee.

It is observed that the ability of the boosted classification tree to make inference about the probability of recession increases as economic indicators are added, especially from model 78s onwards, where production,prices, and interest rates indicators basically come into play. Using the information from these indicators, the estimate of the probability of recession rises to almost one at times classified as recessionary by the Committee. This represents a significant gain in relative efficiency in the classification, compared to the inference that the longer indicators used in the prediction of recessions since 1971. In addition, all four models are able to predict the current recession caused by Covid-19 from the March 2020 data with probability of almost 1.

3.3.2. Pseudo real time analysis

It is not possible to evaluate in real time how the model would perform in predicting 3-month and 6-month ahead recessions in the Spanish economy because we can not obtain real-time indicator data for all the variables included in the model at each point in time at which the predictions would be made. However, we will use an approximation to this exercise by taking subsamples from the longer database and making recursive forecasts with those subsamples.

Let us take as an example of how recursive predictions are constructed for the 3-month prediction exercise of the 71s analysis. To make the first prediction, we start the database in 1974.05, build the classification tree with the Stochastic Gradient Boosting algorithm, and make a prediction of the probability of recession in 1974.08. To make the second prediction in 1974.09, we repeat the operation, but using data up to 1974.06. Following this process iteratively, the last database used is the one truncated in 2020.03 with which the probability of 2020.06 is predicted.

The recursive predictions from the four analyses, 71s, 78s, 88s, and 04s are shown in Figure 3.7, along with the Cycle Dating Committee recessions that are identified by shaded areas. The out-of-sample exercise corroborates the results obtained with the in-sample analysis. As a first remark, it is important to highlight the high ability of the classification trees to predict recession probabilities that rise from almost zero to almost one in recessions in the Spanish economy. The ability to classify the business cycle increases with the 88s model when the information contained in the confidence, stock market, and interest rate indicators is added. To anticipate the Great Recession, prices, financial and construction indicators took a leading role.

As the out-of-sample analysis involves a recursive estimation of the model, we can analyze

the evolution of the relative importance of the indicators. Figure 3.8 depicts this information with some of the most important variables in terms of their highest relative importance. In particular, it represents the relative imporances for the trend-related leading indicator of GDP, car sales leading indicator, registered unemployment, housing construction, the future tendency production indicator from the business tendency survey, the consumer price index, and the monetary market spread with the 78s data set. As in previous cases, recessions in the Cycle Dating Committee are represented with shaded areas. Although the relative importance at the end of the sample is closer for all indicators, the importance differs greatly in its evolution over this period. At the beginning of the sample, consumer price index, employment, car sales and GDP indicators were more important. However, in the mid-1990s, the spread becomes increasingly important and they become the most important leading indicator of the Great Recession together with the consumer price index, while construction indicators during the Great Recession became one of the main coincident indicators. With regard to the recession caused by Covid-19, the GDP, and production indicators seems to increased, opposed to the financial or construction-related ones, since it is a recession caused by the production, due to the forced confinement of economic agents, although not much can be said yet due to data availability[4].

3.4 — Conclusions

The analysis of the business cycle in Spain has gained interest in the scientific literature after the last two recessions, the Great Recession and the crisis derived from covid-19, due to their major economic consequences, in terms of employment, production and, in general, evidenced shortcomings in the Spanish economy. In addition, the availability of more complete databases allows the use of more complex techniques for the analysis.

In this chapter we present boosting regression trees as an alternative capable of synthesizing the information contained in large databases of economic indicators to anticipate recessions in the Spanish economy. For a sample of 270 indicators comprised between 1971.01 and 2020.03, the recursive predictions of recession probabilities 3 and 6 months in advance are almost zero except for the moments identified by the Spanish Cycle Dating Committee, where they rise to almost one.

Using the partial dependence charts we have identified which economic indicators are most influential in predicting recessions in Spain. Over the whole period, among the best indicators for predicting recessions are the leading trend indicators of GDP and car sales, consumer price indexes, and registered unemployment. The ability to anticipate recessions increases

[4]During the recession of 1992 the producer price index had a sharp increase, to later dissapear. It should be mention that including the Labour Force Survey, the unemployment indicator increases systematically while maintaining the dynamic.

when confidence, stock market and interest rate indicators are added. In the Great Recession, construction indicators and some financial indicators related to deposit holdings gained prominence. In the recession generated by Covid-19, indicators of production, unemployment and consumption were once again the most important indicators, since it was a production and consumption crisis rather than a financial crisis.

The approach is not free of limitations and raises future lines of action. In particular, in the applied methodology , temporal dynamics does not play a key role in the elaboration of the tree, being entirely dependent on the richness of the database. In this sense, an alternative would be to analyze the performance and conclusions obtained from incorporating dynamics in the creation of the tree, as is taken into account in autoregressive decision trees (Meek, Chickering, and Heckerman, 2002). Furthermore, to try to lessen the problem of imbalance in the series that we want to classify in the database, expansions and recessions, one could try to adapt some of the recent transfer learning algorithms such as SER or STRUT (Segev et al., 2016).

3.5 — References

Breiman, L., Friedman, J., Olshen, R., and Stone, C., 1984. Classification and regression trees. Chapman and Hall, Wadsworth, New York.

Camacho, M., and Pérez Quirós, G., 2011. Spain-STING: Spain Short Term INdicator of Growth. *The Manchester School* 79: 594-616.

Camacho, M., and Doménech, R., 2012. MICA-BBVA a factor model of economic and financial indicators for short-term GDP forecasting. SERIEs: *Journal of the Spanish Economic Association* 3: 475-497.

Camacho, M., Pacce, M., and Ulloa, C., 2018. Regional business cycle phases in Spain. *Estudios de Economia Aplicada* 36: 875-896.

Camacho, M., Gadea-Rivas, M., and Pérez Quirós, G., 2019. The Great Recession. Worse than ever? *Revista de Economia Aplicada* 76: 73:100

Camacho, M., Caro, A., and López-Buenache,G., 2020. The two-speed Europe in business cycle synchronization. *Empirical Economics*, 59(3), 1069-1084.

Cancelo J. 2004. El ciclo del empleo en España. Un análisis regional. *Revista económica de Castilla La Mancha* 4: 107-138.

Cuevas, A., and Quilis, E., 2012. A factor analysis for the Spanish economy. *Journal of the Spanish Economic Association* 3: 311-338.

Cuevas, A., Pérez Quirós, G., and Quilis, E., 2017. Integrated model of short-term forecasting of the Spanish Economy (Mipred Model). *Revista de Economía Aplicada* 25: 5-25.

Döpke, J., Fritsche, U., and Pierdzioch, C., 2017. Predicting recessions with boosted regression trees. *International Journal of Forecasting* 33: 745-759.

Freund, Y., 1995. Boosting a weak learning algorithm by majority. *Information and Computation* 121: 256-285.

Freund, Y., and Schapire, R., 1996. Experiments with a new boosting algorithm. *Proceedings of ICML* 13: 148-156.

Friedman, J., Hastie, T., and Tibshirani, R., 2000. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics* 28: 337-407.

Friedman, J., 2001. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* 29: 1189-1232.

Friedman, J., 2002. Stochastic gradient boosting. *Computational Statistics and Data Analysis* 38: 367-378.

Gadea-Rivas, M., Gomez-Loscos, A., and Montañes, A., 2012. Cycles inside cycles: Spanish regional aggregation. *Journal of the Spanish Economic Association* 3: 423-456.

Gadea-Rivas, M., Gomez-Loscos, A., and Bandrés, E., 2018. Clustering regional business cycles. *Economic Letters* 162: 171-176.

Gadea-Rivas, M., Gomez-Loscos, A., and Leiva-León, D., 2019. Increasing linkages among European regions. The role of sectoral composition. *Economic Modelling* 80: 222-243.

Hastie, T., Hastie, T., Tibshirani, R., and Friedman, J., 2001. The elements of statistical learning: Data mining, inference, and prediction. New York: Springer.

IMF, 2009. World Economic Outlook-April 2009: Crisis and Recovery.

Mease, D., Wyner, A., and Buja, A., 2007. Cost-weighted boosting with jittering and over/undersampling: Jous-boost. *Journal of Machine Learning Research* 8: 409-439.

Meek, C., Chickering, D. M., and Heckerman, D., 2002.. Autoregressive tree models for time-series analysis. In Proceedings of the 2002 SIAM International Conference on Data Mining (pp. 229-244). Society for Industrial and Applied Mathematics.

Ng, S., 2014. Boosting recessions. *Canadian Journal of Economics* 47: 1-34.

Páez, A., López, F., Ruiz, M., and Camacho, M., 2019. Inducing non-orthogonal and non-linear decision boundaries in decision trees via interactive basis functions. *Expert Systems with Applications* 122: 183-206.

Piger, J., 2020. Turning points and classification. In Macroeconomic forecasting in the era of big data: Theory and application, Peter Fuleky (editor). Springer International Publishing.

Schapire, R., 1990. The strength of weak learnability. *Machine Learning*, 5: 197-227.

Schapire, R., and Freund, Y. 2012. Boosting: Foundations and algorithms. MIT Press.

Segev, N., Harel, M., Mannor, S., Crammer, K., and El-Yaniv, R., 2016. Learn on source, refine on target: A model transfer learning framework with random forests. IEEE transactions on pattern analysis and machine intelligence, 39(9), 1811-1824.

Ward, 2017. Spotting the danger zone: Forecasting financial crisis with classification tree ensembles and many predictors. *Journal of Applied Econometrics* 32: 359-378.

3.6 — Appendix A

Alternative applications of relative importance

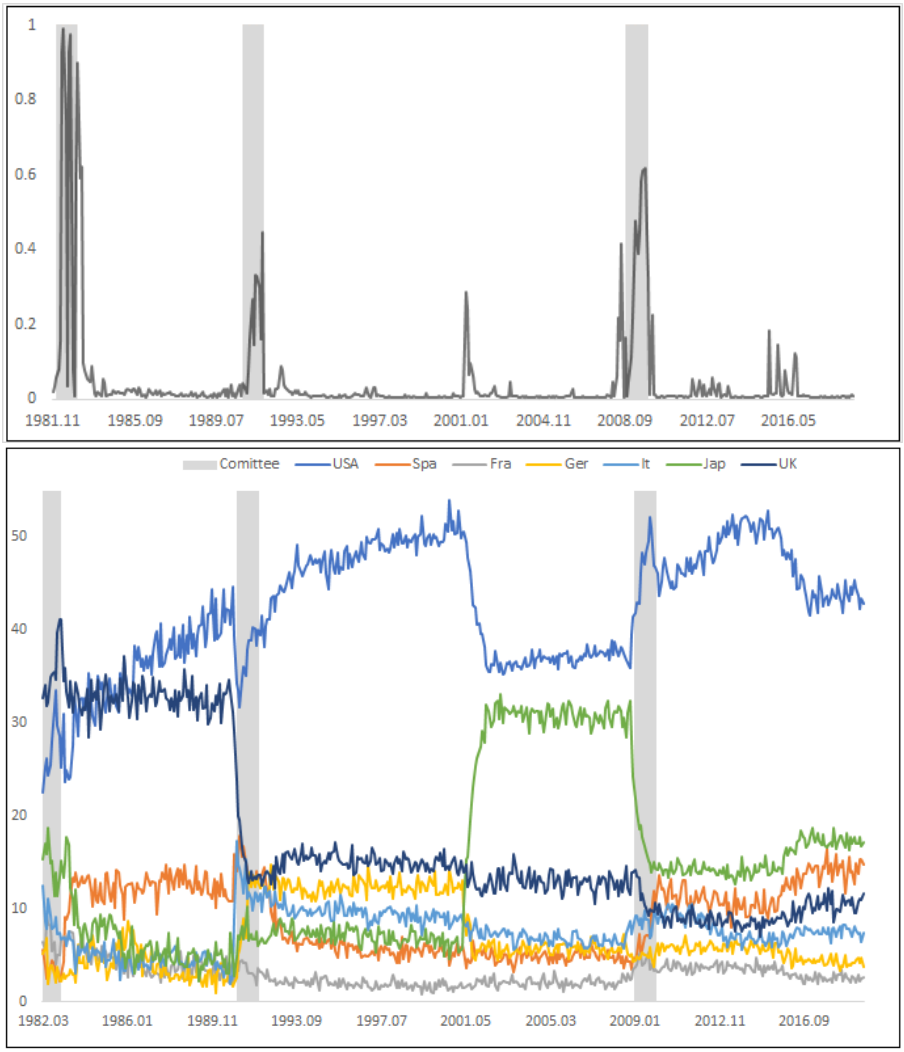
Relative importance is a very useful tool in the case of a sufficiently accurate prediction of the probability of recession. In addition to the analysis of variables carried out for the case of Spain in the main part of the chapter, this appendix shows two other possible applications of this tool.

First, the use of relative importances in our context can help us also to analyze which countries are the most helpful in predicting the probability of a global recession. If we get a good prediction of global recessions in terms of accuracy, through the out-of-sample exercise we would obtain information on which country at each point in time was key to predicting an adverse economic situation, or which country is most helpful in establishing that the probability of global recession is low.

As an example, we use a sample of seven countries such as the United States, Japan, United Kingdom, Germany, France, Spain and Italy to predict global recessions, and thus to analyze which countries helped at each period to make the prediction of recession in the world economic cycle. We use the IMF's criteria to define a global recession as a decrease in the yearly per capita real global Gross Domestic Product accompanied by a deterioration in at least one of the next seven additional global macroeconomic indicators: industrial production, international trade, capital movement, oil consumption, unemployment rate, per capita investment, or per capita consumption (IMF, 2009). On the one hand, Figure A1 shows in the upper panel an adequate prediction, although it does not reach a maximum level of probability of recession. On the other hand, from the lower panel it can be noted how initially UK and US were the most relevant in terms of importance to make the prediction, although after the 1991 recession

the US is dominant for it, with a certain decline around 2001 (dotcom crisis) when Japan gains weight, to recover it with the Great Recession, loosing some importance from 2015, and gaining some Spain and UK.

Figure A1. Out-of-sample probability of worldwide recession and country relative importances



Notes. Panel A. Probability of recession estimated with Production indexes from USA, Japan, UK, Germany, France, Italy, and Spain. Shaded areas identify recessions according to IMF definition. Panel B. Relative importances for each of the countries to produce probabilities.

Second, through relative importance we can also create a measure of synchronization between two economies. In particular, let us use the same variables, representative of the economy, to create the recession probabilities for two countries. If a similar recession probability is indicated in both cases, and, moreover, the relative importance of the different variables in creating the recession probability are similar in both cases, both economies will have a degree of synchronization both temporally and structurally.

Therefore, we will be able to simply create a measure of temporal and structural synchronization by measuring the distance between the recession predictions and a measure of

similarity of the two vectors of relative importances, which can be measured in time through the out-of-sample estimation. The latter can be obtained by a similarity measure, such as the cosine similarity, which is defined as the inner (or scalar) product of two vectors,

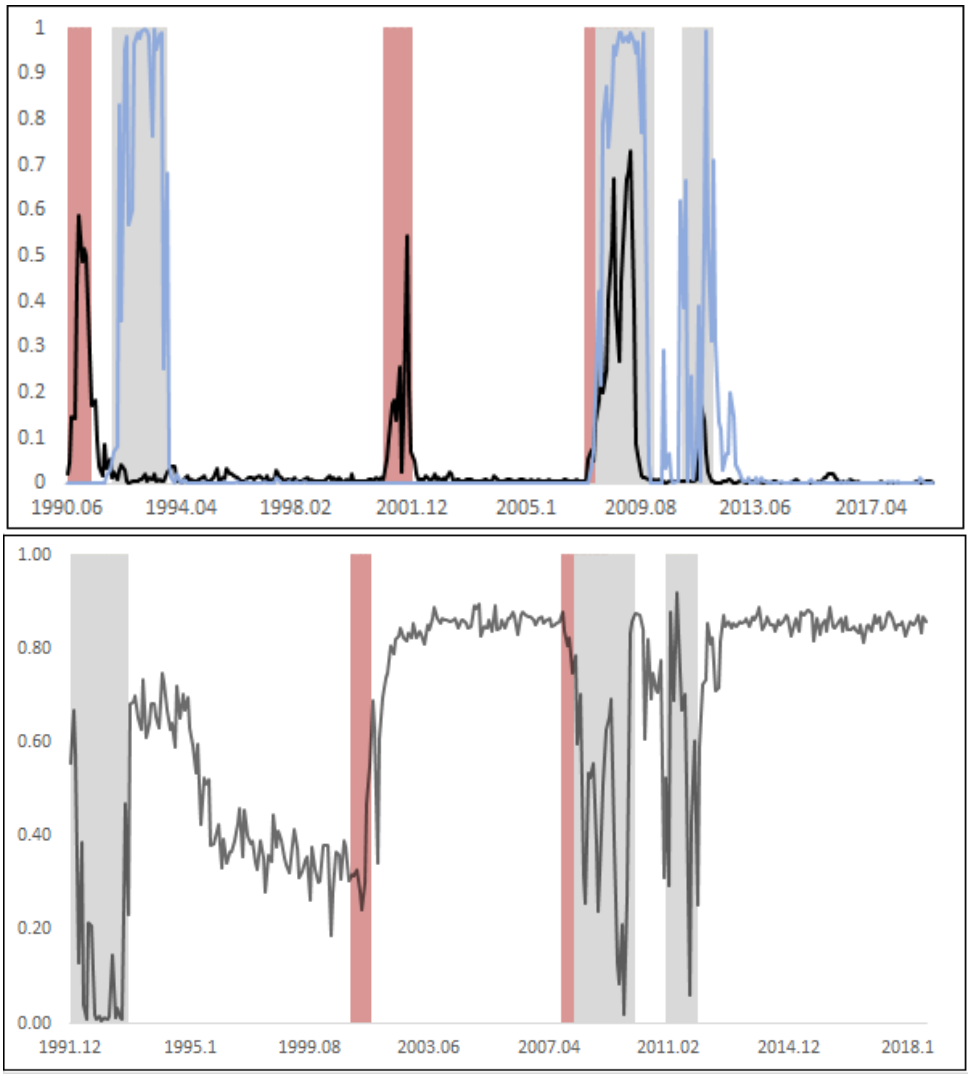
$$cosine_{ij,t} = \frac{\sum_{n=1}^{N} IB_{i_n,t} IB_{j_n,t}}{\sqrt{\sum_{n=1}^{5} IB_{i_n,t}^2 \sum_{n=1}^{5} IB_{j_n,t}^2}},$$

where n represents the N relative importances at time t of the N variables used to create the probability of recession, and i, j the respective countries. Thus, the synchronization will be estimated as

$$sync_{ij,t} = (1 - |p_i^R - p_j^R|) \cdot cosine_{ij,t},$$

for $t = 1, ..., T - h$. In particular, we gathered eight variables for both United States and Spain, which are available at least from 1986, from the st. Louis Fred database: production index, consumer price index, total share prices, OECD composite indicator of confidence from consumer opinion survey, 10-year goverment bond yields, composite indicator of confidence from business tendency surveys for manufacturing, unemployment rate, and spread from 10-year treasury minus federal funds rate. All variables where used in year over year terms. Figure A2 depicts the predicted recessions probabilities and the synchronicity measure together with the recession from NBER and Spanish Business Cycle Dating Committee. It can be noted that the predictions match the recessions and besides, that in periods of expansions the synchronicity measure tends to be larger while in recessionary periods uses to fluctuate.

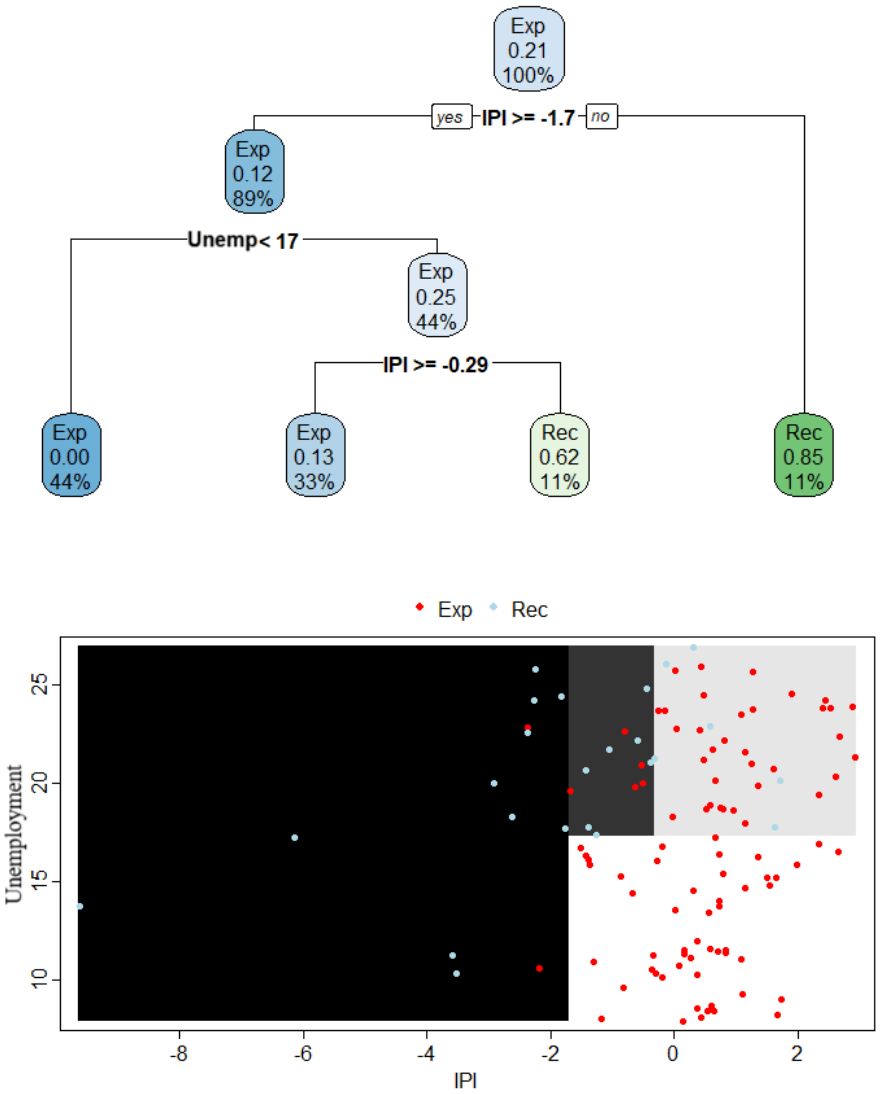
Figure A2. Out-of-sample synchronicity for US and Spain



Notes. Panel A. Probability of recession estimated with the same indicators for USA (black line) and Spain (light blue line). Shaded areas in grey identify Cycle Dating Committee recessions and shaded areas in light red identify NBER recessions. Panel B. Synchronicity measure between US and Spain.Shaded areas in grey identify Cycle Dating Committee recessions and shaded areas in light red identify NBER recessions.

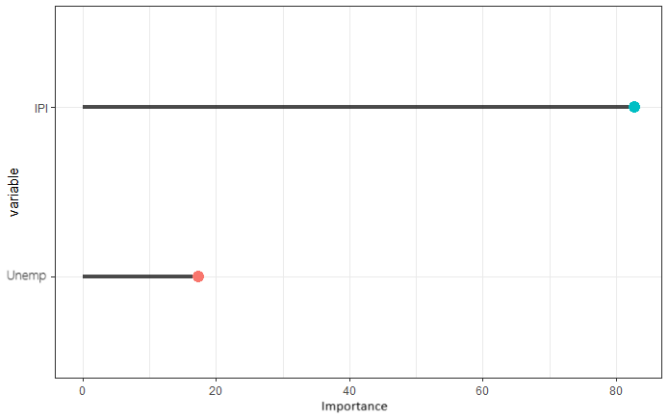
3.7 — Figures

Figure 3.1. Classification tree example



Notes. Classification tree analysis applied to the quarterly aggregation of industrial production and registered unemployment data (EPA survey). In the upper graph, each node shows the classification, the percentage of recessions and the percentage of recessions observations. In the lower graph the shaded areas are darker the higher the estimated probability of recession.

Figure 3.2. Relative importance example.

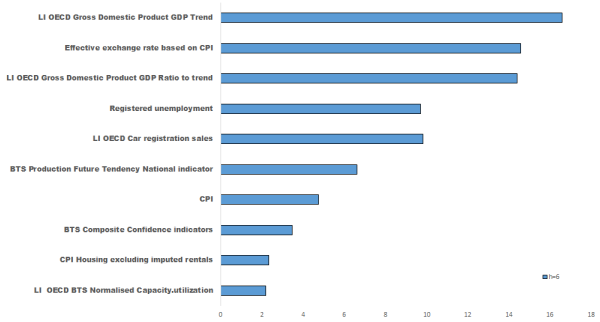


Notes. Variables are the quarterly aggregation of industrial production and the registered unemployment data from EPA survey.

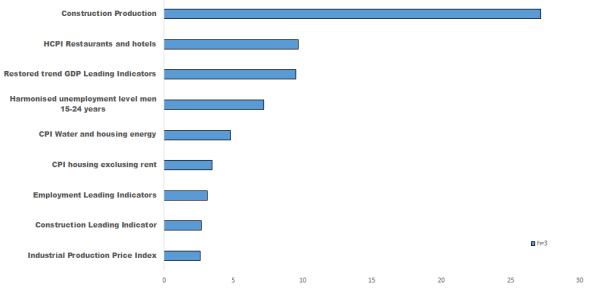
Figure 3.3. In-sample relative importance

(a) Panel A. Sample from 1971. h=3

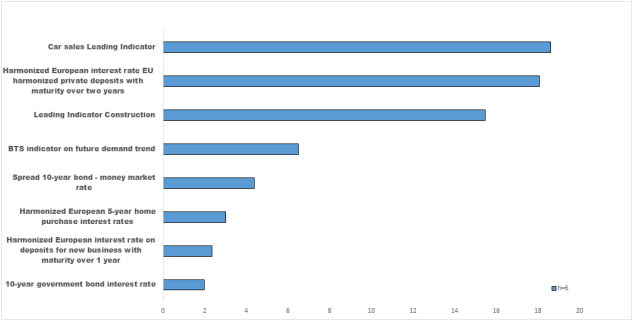
(b) Panel B. Sample from 1971. h=6



(c) Panel C. Sample from 2004. h=3

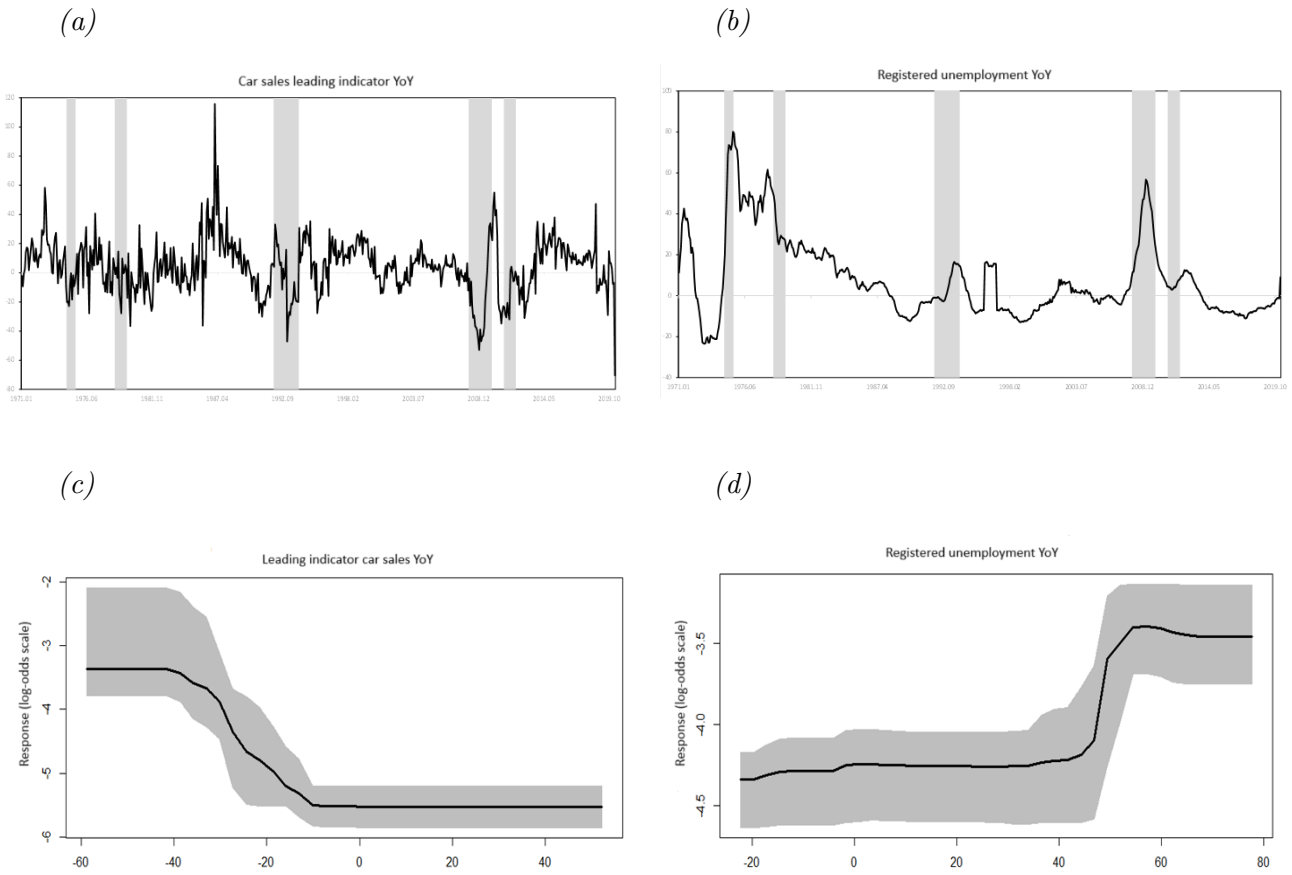


(d) Panel D. Sample from 2004. h=6



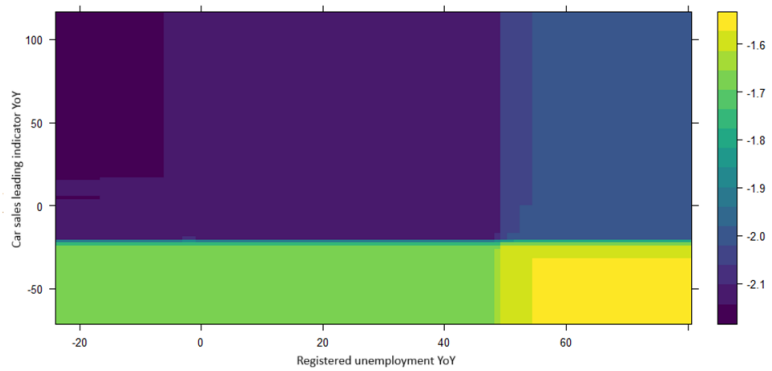
Notes. Relative importance have been scaled so that they add up to 100.

Figure 3.4. Partial dependence



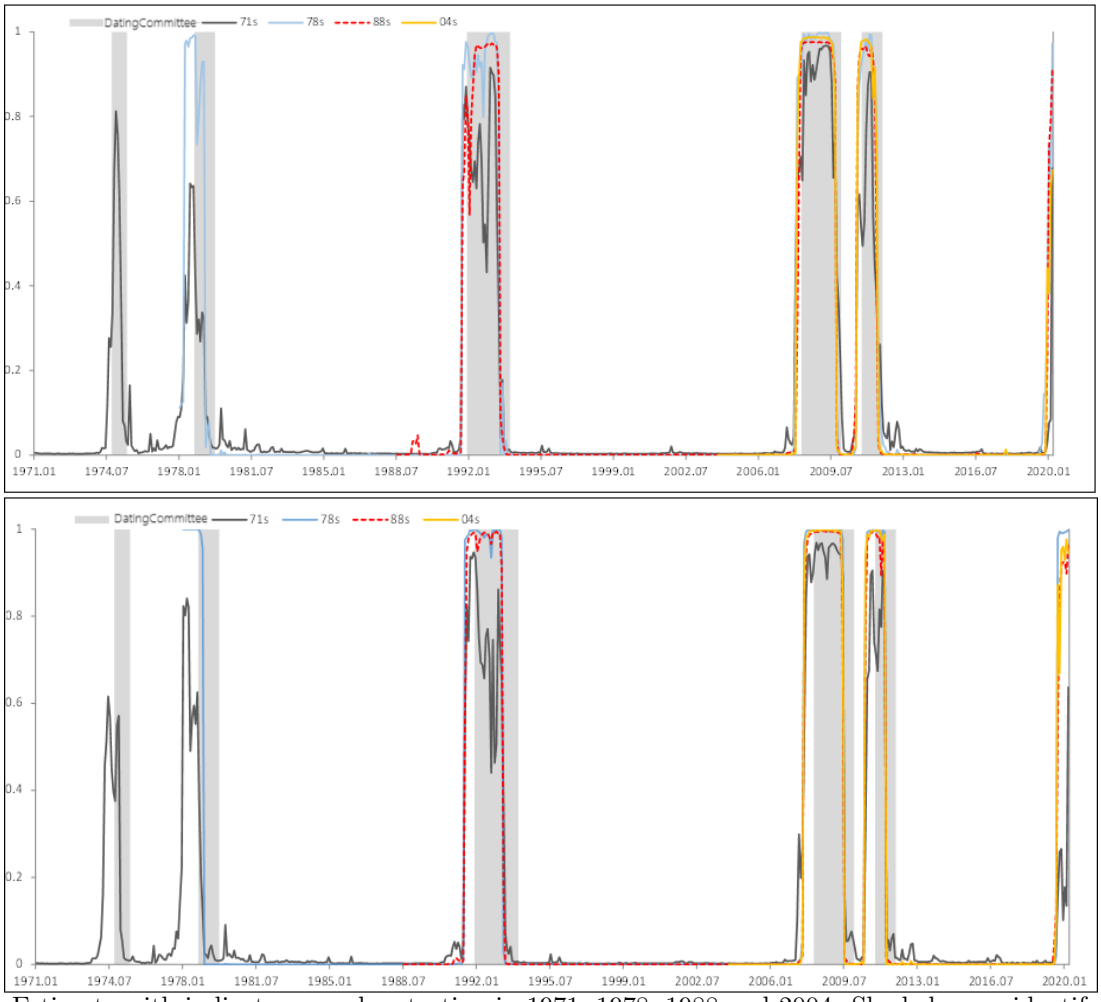
Notes. Top charts: year-on-year growth rates of the leading indicator of car sales and registered unemployment; shaded areas are recessions according to the Cycle Dating Committee. Lower graphs: effect on the logarithm of the 6-month odds ratio; shaded areas show 95% confidence intervals. Sample period: 1971.01-2020.03.

Figure 3.5. Interaction effect.



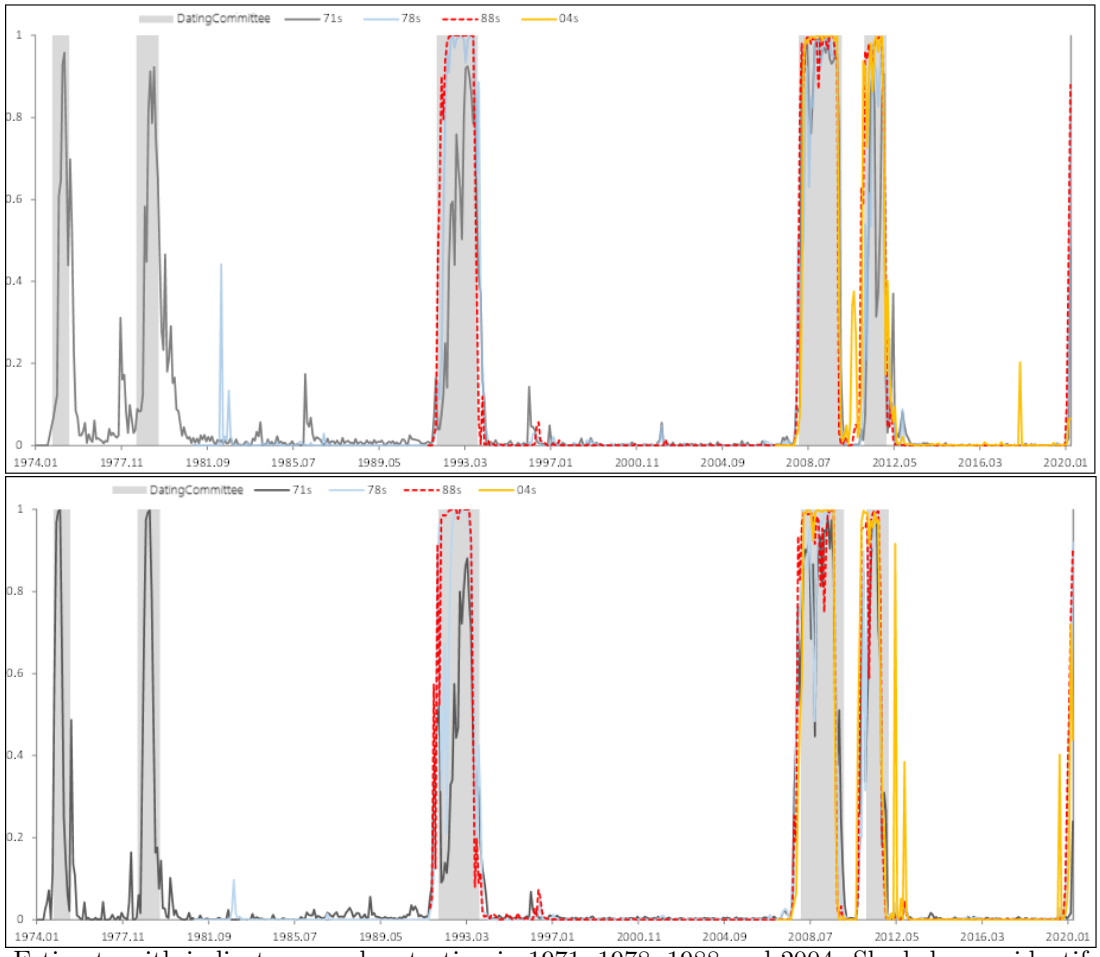
Notes. Joint effect on the logarithm of the 6-month odss ratio of the year-on-year growth rates of the leading indicator of car sales and the year-on-year increase in registered unemployment.

Figure 3.6. In-sample probability of recession



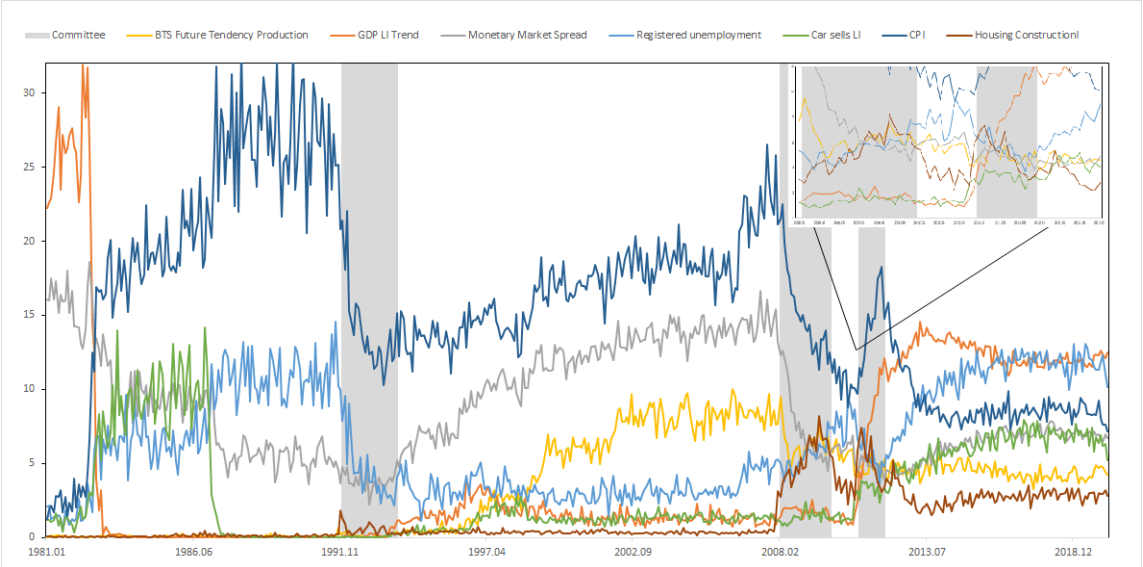
Notes. Estimate with indicator samples starting in 1971, 1978, 1988 and 2004. Shaded areas identify Cycle Dating Committee recessions.

Figure 3.7. Out-of-sample probability of recession



Notes. Estimate with indicator samples starting in 1971, 1978, 1988 and 2004. Shaded areas identify Cycle Dating Committee recessions.

Figure 3.8. Relative importance evolution.



Notes. Results obtained with the database since 1978. The shaded areas make reference to the recessions of the Cycle Dating Committee. A zoom representation from 2008 to 2012 is presented in the upper right corner.

Chapter 4

Predictive ability of a dynamic factor model predicting homicides with gun

4.1 — Introduction

The tracking of crime levels is valuable for many economic sectors, given its potential impact on decision making and risk management. For instance, insurance companies assess the risk associated with their clients' locations by assigning higher premiums in areas with higher crime levels (Barr and Pease, 1990), whereas real estate agents consider crime levels while pricing properties (DeLisle, Never and Grisson, 2020). In addition, tourism industry benefits from safe destinations (Altindag, 2014), which might boost for example local economies. Thus, crime data analysis is essential not only for public policy design and for its impact on economic growth (Estrada and Ndoma, 2014), but also for resource allocation and strategic planning in multiple economic sectors.

Despite years of drastic reductions in crime levels, certain types of violent crimes have recently increased in the US (Rosenfeld and Lopez, 2020). Gun violence is especially on the rise, as suggested by Albrecht (2022). Lately, we are witnessing a steady increase in non-fatal crimes, as documented by the Police Executive Research Forum (Police Executive Research Forum, 2021) and supported by the Giffords Law Center (Giffords Law Center, 2022). Just as non-fatal shootings are spiking across several cities in the country, fatal shootings are rising, as noted by the Pew Research Center (Gramlich, 2022) from data on gun death rates. This dramatic trend has attracted media attention and raised concerns in Americans about gun violence. The 2022 Election Tracking Survey from Ipsos (IPSOS, 2022) identifies gun violence as the second most important issue for Americans.

As an example of this recent trend, Figure 4.1 shows the 10-year relative increase in homicide offenses rates for different weapons registered on the National Incident-Based Reporting System (NIBRS) (Bureau of Justice Statistics, 2022). While those performed with firearm have dramatically increased, assaults with knives (jackknifes, cutting instruments, etc.), personal weapons (hands, feet, etc.), and other weapons (blunt objects, motor vehicles, explosives,

etc.) have remained at the same level, or even decreased, in the last ten years (Rosenfeld and Fox, 2019).

The ability of policymakers to efficiently make decisions that can reduce crime relies on wide and quick access to real-time data, as suggested, for instance, in Neumayer (2003), Drake et al. (2009), and Barreras et al. (2016). Evidence-based legislation is being embraced throughout the world. In the US, the Foundations for Evidence-Based Policymaking Act of 2018 became a law in 2019 (Congress.Gov, 2019), thereby requiring public access to agencies' data[1] and use of statistical evidence for any bill. Likewise, the Organisation for Economic Co-operation and Development points that countries should coordinate their strategies for policy evaluation with those related to evidence and data governance (OECD, 2020). However, several pitfalls persist with respect to availability, quality, and saliency of firearm-related violence data in the US (Strom and Smith, 2017; Roman and Cook, 2021).

According to the July 2014 National Criminal Justice Report from the Bureau of Justice Statistics (Regoeczi et al., 2014), the Wonder database of the National Center for Health Statistics from the Centers for Disease Control and Prevention (CDC) and the Supplementary Homicide Report (SHR) from Federal Bureau of Investigation (FBI) are the only source of detailed information on homicides (Centers for Disease Control and Prevention, 2021) While SHR is published with an annual frequency and suffers from missing or incomplete reports (Fox and Swatt, 2009), the CDC database contains granular data at a monthly resolution. Such an improvement in data quality comes at the expense of a significant delay in its release, as acknowledged by Mancik et al. (2021). Specifically, in the CDC database all deaths for each month of the year are released all at once at the end of the following year. In this way, January data is released on December of the following year (23 months later) and December data is released on December of the following year (12 months later). The large amounts of data that need to be collected and processed are partly the reason why this data is published with such a hold-back. As a result, the actual picture of the homicides with firearm in the US can have a delay between 12 to 23 months, depending on the month of the analysis. This time delay cannot be compensated by using other series that are released with shorter delays without losing relevant information, since only 40% of violent crimes are reported to the police (Morgan and Thompson, 2021).

Mathematical tools could be used to overcome some of the practical drawbacks in data access, thereby supporting evidence-based interventions. However, to date, there is a paucity of reliable tools to gauge firearm violence in the US. Several efforts from the scientific community have tackled the problem of modeling temporal dynamics of crime-related rates, but these methods have seldom been applied to homicides and almost never to homicides with firearms – the objective of this chapter. Back in 2003, Gorr et al. (2003) already proposed a Holt exponential model to forecast crime series, which, however, is questionable when attempting

[1] An agency is defined under section 901(b) of title 31.

long-term predictions (Chatfield and Yar, 1988; Alonso-Brito et al., 2021) and is fragile for its excessive reliance on data extrapolation (Gardner and McKenzie, 1985). Since then, a range of methods based on time-series analysis have been adapted to study crime rates in general (Berk, 2008), including homicides (Phillips, 2016).

Recently, Feng et al. (2018) has suggested that, among artificial intelligence (AI) techniques, tree models may outperform other approaches, such as k-nearest neighbors or naive Bayes for the prediction of crime, an observation which is in line with the previous findings by Nasridinov et al. (2013). In this vein, Berk et al. (2009) used random forests to specifically forecast homicides of paroles within two years after intake. Meskela et al. (2020) and Devi and Kavitha (2021) have proposed the use of a specific type of neural network particularly useful in time-series modeling, the long short-term memory recurrent neural network, to automate crime prediction. While promising, AI techniques are data hungry, whereby their performance is controlled by the richness of their training dataset (Carvalho et al., 2018), so that overfitting is difficult to avoid (Vezhnevets and Barinova, 2007).

In parallel to implementations of AI techniques, Cesario et al. (2016) and Yadav and Sheoran (2018) have used autoregressive integrated moving average (ARIMA) models to predict crime rates. These models are known to systematically revert the forecasts to the mean of the series (Deadman et al. 2001), making it difficult create reliable predictions. Building on the classical ARIMA, researchers have examined the percent change techniques (McDowall, 2002), ARIMA with fan charts (Yim et al., 2020), spatio-temporal autoregressive models (Shoesmith, 2013), and generalized least squares regression to study homicides.

Beyond univariate autoregressive models, classical multivariate autoregressive approaches were also adopted to study crime (Blumstein and Rosenfeld, 2008). The ability to account for multiple drivers in a multivariate sense allows for capturing salient phenomena that would be otherwise missed. For instance, one of the argued reasons for the registered increase in violence with firearm from 2014 to 2018 is the "Ferguson effect" (Hoffman et al., 2021; Cheng and Long, 2022). According to this theory, the police has been more scrutinized following the protests in Ferguson in 2014 in the wake of Michael Brown's death, thereby changing their approach to law enforcement, and, in turn, to crime prevention. Likewise, the number of active law enforcement officers has been suggested to drive the dynamics of crimes (Parker et al., 2017), along with economic fluctuations (Rosenfeld and Fornango, 2007) and firearm possession (Cook and Ludwig, 2006).

In this vein, Pratt and Lowenkamp (2002) related homicides to time-series of coincident economic indicators through a bivariate ARIMA model, whereas Cherian and Dawson (2015) has employed a vector autoregressive (VAR) model to predict several category crimes and Parkin et al. (2020) studied, within the same approach, the relation between deadly force incidents, line of duty deaths, and homicides rates. Although offering a much more complete view of homicide dynamics, most of the series that may be useful for multivariate time-series

analysis could have very different sample periods and sampling frequency (so-called "ragged edge" problem identified by Wallis (1986)). Bringing the series to the same sampling period and frequency might cause an excessive reduction in the dataset, thereby hindering the use of any VAR model.

To bestow improved prediction of homicides, we relied on a single-index dynamic factor model (DFM) approach (Sargent et al., 1977; Sotck and Watson, 1991), which allows for the integration of information from multiple, easily accessible time-series to predict a variable of interest. Single-index DFM relies on the co-movement of different series, thereby reducing dimensionality compared to a VAR approach and improving the reliability of the identification. The DFM framework allows for the use of data with different frequencies, as shown in Harvey (1990) and Mariano and Murasawa (2003), to create common and idiosyncratic dynamics. By using Kalman filtering to fill any missing observations (Brockwell and Davis, 2009), DFM forecasts benefit from the different delays of the time-series. The proposed approach solves some of the limitations of the existing state-of-the-art. First, DFM is less prone to overfitting than AI techniques when working with real datasets of crimes that have limited size and potentially missing data (Mitchell and Mitchell, 1997; Ying, 2019; Soybilgen, 2020). Second, the identification of common and idiosyncratic dynamics within DFM avoids the problems of autoreregressive models related to the reversion to the mean (Beshears et al., 2013; Nau, 2014) and the inability to detect abnormal periods (Stock, 1994; Wheeler and Kovandzic, 2018; Yim et al., 2020).

Through the proposed DFM, we are able to predict homicides with firearm in the short term better than other benchmark models, such as tree-based models, neural networks, and classical autoregressive approaches. Specifically, through an out-of-sample exercise, we find that the DFM is the only approach that can perform a better prediction than a benchmark ARIMA model, on average for every month in the year. Furthermore, the model shows an improved ability to timely and accurately capture unexpected changes in the direction of the series, as those experienced in the recent COVID-19 outbreak. Not only does the model offer improved qualitative agreement with real data, but also it begets higher predictive accuracy. The enhanced ability to predict homicides with firearm offers a vantage point to policymakers and practitioners, allowing for timely predictions that would be otherwise unfeasible.

The rest of the chapter is organized as follows. Section 2 details a preliminary analysis of the data and depicts the the methodology of the dynamic factor model. Section 4 applies the approach to describe the dynamics of the homicides with guns, and analyzes the accuracy of the nowcast and forecast ability. This section also compares the estimation from the model with machine learning techniques. Section 4 concludes and outlines some further research lines. This is followed by the references, appendices, tables and figures referred to in the chapter.

4.2 — Data and methods

Data

Homicides with firearm was our main variable of interest. On the Wonder database, we collected data on deaths whose "injury intent" was "homicide" and whose "injury mechanism and all other leading causes" were "firearm" from January 1999 to December 2020. This data is updated by the CDC once a year at a monthly resolution, and it is based on death certificates of US residents from the National Center of Health Statistics. Since there is no official monthly population data, we used the monthly population estimates (Bureau of Economic Analysis, 2022) from the Economic Bureau of Analysis to compute per-capita homicides with firearm.

Our approach is based on the use of a single-index dynamic factor model, which seeks to unveil a general co-movement in the data. To construct a common dynamics in the dynamic factor model, we should consider homicide-related data with a certain degree of co-linearity with homicides with firearm. The model's ability to deal with missing observations allows for taking into account data with different sampling periods. In this vein, the integration of series whose data is released before the official release of data about homicides with firearm is expected to improve predictions. With this in mind, we collected the following additional variables:

- **A provisional estimation of homicides with firearm**. Since January 2017, the CDC publishes a provisional quarterly estimation of the rate of homicides with firearm, released ten months after the end of the quarter (that is, a cumulative monthly series with two missing observations). The rate for the third quarter of 2020 was the highest recorded in our observation window ending in December 2020, with an annualized rate of 15 homicides with firearm per 100,000 inhabitants.

- **Data on deaths in incidents involving guns**. From January 2014, the Guns Violence Archive (Guns Violence Archive, 2022), a non-profit corporation, registers daily gun violence incidents from law enforcement, media, and commercial sources. Based on the incidents, they report the victims of gun violence, including murders, accidents, or suicides. We specified the filter selection "Shot-Dead (murder, accidental, suicide)" as "Incident Characteristic" to create a monthly series of deaths in incidents with guns from January 2014 to December 2020. The highest number of deaths was recorded on July 2020, with 1,964 deaths.

- **Data on homicides from three of the main cities in the US**. We collected daily crime rates in New York City, Chicago, and Philadelphia from the police departments. These data are publicly available from January 2006, weekly updated, and registered

within the Unified Crime Report guidelines from the FBI. The Census Bureau Population data for each city was used to estimate the per-capita rates, which were further averaged among each other to create a baseline estimate. In our observation window from January 2006 to Decemeber 2020, the highest average rate was recorded on July 2020, with 2.41 homicides per 100,000 inhabitants.

- **Media output**. According to Phillips and Hensley (1984), when publicity is given in mass media to violence events, an increase in mortality is likely to follow. We collected media output from January 1999 to December 2020, using the Proquest Database, by searching for the number of news articles containing the words "homicide" and "shot" from the New York Times (2,997 news articles) and the Washington Post (6,175 news articles). Daily data was aggregated to create a monthly time-series where the highest media output was recorded on December 2020, with 107 news articles. Similarly, we collected the number of news containing the word "riots" (24,801 news articles) and those containing the word "unemployment" (6,925 news articles). The highest media output on riots was recorded in June 2020, when 744 news articles were published, whereas the peak of news articles about unemployment was in July 2020, with 195 news articles registered.

- **Google Trends**. Building on recent work on crime analysis with Google Trends (Gamma et al., 2016), we collected data from Google Trends with the search term "homicide" from January 2004 to December 2020 at a monthly resolution. The highest output was recorded on January 2013 (index of 100). We repeated the procedure with the term "gun", for which the highest output was observed on December 2012.

- **Background checks**. This classical proxy of firearm sales (Lang, 2013; Wallace, 2015) is made available by the FBI at a monthly resolution via the National Instant Criminal Background Check System (NCIS) (FBI, 2022a). In our observation window from January 1999 to December 2020, the highest number of background checks was recorded on December 2020, when 3,937,066 background checks were performed throughout the nation.

- **Economic uncertainty**. The Economic Policy Uncertainty (EPU) index, available at a monthly resolution from January 1999 from Baker et al. (2016), was chosen as a proxy of economic uncertainty. In our observation window from January 1999 to December 2020, the highest score was registered in May 2020, with an index of 350.

- **Microblogging data**. Building on recent work (Chen et al., 2015) on the use of Twitter data in modeling crime, we collected the number of daily geo-located tweets in the US containing the word "homicide" from January 2010 to December 2020. We collected a total of 146,661 posts and created a monthly time-series, whose peak was registered on May 2015 with 1,862 tweets.

For all the series mentioned except of the media output, a monthly seasonal adjustment was performed to remove seasonal effects. Besides, all the series except of homicides with firearm, monthly and provisional, and deaths in incidents with guns, were also detrended. The resulting time-series are plotted in Figures 4.2a to 4.2j. All the raw and processed data are available at Github.

The data differ in their resolution and release date throughout the year (Table 1). While variables such as background checks and media output are available in the same time-span of homicides with firearm (starting January 1999), Google Trends' data is available only from 2004 and homicides data in the three cities from 2006. All the variable, except of homicides with firearm and their provisional estimates, are released with no delay. Firearm homicides have a delay of 12 up to 23 months; for example data from January 2020 to December 2020 of homicides with firearm was released in December 2021. The provisional quarterly estimation is released with a fixed ten months of delay. Other variables available at yearly resolutions and released with similar delays as homicides with firearm (number of police officers (FBI, 2019), the National Crime Victimization Survey (Bureau of Justice Statistics, 2017), or the FBI violent crime index (FBI, 2019)), were not taken into account as they would not improve predictive power. Likewise, the registered murders in the recently released Quarterly Tables from NIBRS (FBI, 2022b) were not used due to its short historic period (three editions released up to date).

Dynamic factor model

Single-index dynamic factor models decompose the dynamics of the chosen observable variables $y_{i,t}$, for $i = 1, ..., n$ and $t = 1, ..., T$, as the sum of two unobservable and orthogonal components: one affecting all the time-series, f_t, and the other one accounting for their idiosyncratic variation, $u_{i,t}$. More specifically, we write

$$y_{i,t} = b_i f_t + u_{i,t}, \tag{C.1}$$

where b_i is the loading factor of each variable on the common factor[2]. Any variable at a coarser resolution than the variable of interest (like the provisional estimate of homicides with firearm in Table 1) can be written as the sum of unobserved variables at the chosen resolution with proper delays[3]. The dynamics of the common factor and idiosyncratic components are described as

[2]Other more complex specifications can be accommodated. For example, Guerrón-Quintana, Khazanov, and zhong, (2021) propose that the factor be composed by the addition of two terms, one based on first-order derivatives and the second based on more complex second order derivatives, with time-varying volatility and an interaction between the factor and the factor's innovation. However, this requires a more complex estimation by means of an Unscented Kalman Filter, and in our case, with a simpler and more limited database, its convergence would be unlikely to be achive, so we opted for the more standard formulation.

[3]For example, let $i = 2$ be a time-series with a quarterly resolution, then, we would write $y_{2,t} = b_2 (f_t + f_{t-1} + f_{t-2}) + u_{2,t}$.

autoregressive processes of orders p and q, that is,

$$\begin{aligned} f_t &= \beta_1 f_{t-1} + ... + \beta_p f_{t-p} + \epsilon_t^f \\ u_{i,t} &= c_{i,1} u_{i,t-1} + ... + c_{i,q} u_{i,t-q} + \epsilon_{i,t}^u, \end{aligned} \tag{C.2}$$

where ϵ_t^f captures the errors in the common factor and $\epsilon_{i,t}^u$ the errors in the ith idiosyncratic terms. Errors in the common factor and in the ith idiosyncratic term are considered independent and identically distributed (i.i.d.) in cross-section and time, following $\mathcal{N}(0, \sigma^2_{\epsilon^f})$ and $\mathcal{N}(0, \sigma^2_{\epsilon_i^u})$ distributions, respectively, where σ_{ϵ^f} and $\sigma_{\epsilon_i^u}$ are the standard deviations. Since the common factor is not observed either in mean and variance, σ_{ϵ^f} must be chosen to make the model identifiable (in particular, we set it equal to one).

The model in equations (C.1) and (C.2) can be cast in a state-space representation of the form

$$Y_t = H h_t, \tag{C.3}$$

where h_t is the state vector created from the common and idiosyncratic components in time. The transition equation is written as

$$h_t = F h_{t-1} + \epsilon_t, \tag{C.4}$$

with ϵ_t being i.i.d. $\mathcal{N}(0,Q)$ and $Q = \text{diag}(\sigma^2_{\epsilon^f}, 0, 0, \sigma^2_{\epsilon_1^u}, 0, 0, ..., \sigma^2_{\epsilon_n^u}, 0, 0)$.

For example, considering a model with $p, q = 2$ and all variables at a monthly resolution expect of $y_{2,t}$ chosen to be quarterly, equation (C.3) reads

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ . \\ . \\ . \\ y_{n,t} \end{pmatrix} = \begin{pmatrix} b_1 & 0 & 0 & 1 & 0 & 0 & 0 & ... & 0 & 0 & 0 \\ b_2 & b_2 & b_2 & 0 & 0 & 0 & 1 & ... & 0 & 0 & 0 \\ . & & & & & & & & & & \\ . & & & & & & & & & & \\ . & & & & & & & & & & \\ b_n & 0 & 0 & 0 & 0 & 0 & 0 & ... & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ u_{1,t} \\ u_{1,t-1} \\ u_{1,t-2} \\ u_{2,t} \\ . \\ . \\ . \\ u_{n,t} \\ u_{n,t-1} \\ u_{n,t-2} \end{pmatrix}, \tag{C.5}$$

and equation (C.4) becomes

$$\begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ u_{1,t} \\ u_{1,t-1} \\ u_{1,t-2} \\ u_{2,t} \\ . \\ . \\ . \\ u_{n,t} \\ u_{n,t-1} \\ u_{n,t-2} \end{pmatrix} = \begin{pmatrix} \beta_1 & \beta_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{1,1} & c_{1,2} & 0 & 0 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & c_{2,1} & c_{2,2} & 0 & ... & 0 & 0 & 0 \\ . & & & & & & & & & & & & \\ . & & & & & & & & & & & & \\ . & & & & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & c_{n,1} & c_{n,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ u_{1,t-1} \\ u_{1,t-2} \\ u_{1,t-3} \\ u_{2,t-1} \\ . \\ . \\ . \\ u_{n,t-1} \\ u_{n,t-2} \\ u_{n,t-3} \end{pmatrix} + \begin{pmatrix} \epsilon_t^f \\ \epsilon_{t-1}^f \\ \epsilon_{t-2}^f \\ \epsilon_{1,t}^u \\ \epsilon_{1,t-1}^u \\ \epsilon_{1,t-2}^u \\ \epsilon_{2,t}^u \\ . \\ . \\ . \\ \epsilon_{n,t}^u \\ \epsilon_{n,t-1}^u \\ \epsilon_{n,t-2}^u \end{pmatrix}, \quad \text{(C.6)}$$

where, due to the definition of ϵ_t and Q, the error terms of the period $t-1$ will be zero. The terms in these equations could be estimated by maximum likelihood through a Kalman filter.

In the case of missing observations, such as when treating quarterly data within our monthly resolutions, we use the approach given by Mariano and Murasawa (2003). Specifically, the missing values are replaced by random draws, α_t, from a distribution that does not depend on the parameter space that characterizes the filter, for example $\mathcal{N}(0, \sigma_\alpha^2)$.

The estimation algorithm for obtaining the parameters can be summarized as follows. Let $h_{t|\tau}$ be the estimate of h_t with information up to period τ, that is, the expected value of the state vector conditioned on the past. Denoting $P_{t|\tau}$ its covariance matrix, the prediction equations for the Kalman filter are

$$h_{t|t-1} = F h_{t-1|t-1}, \quad \text{(C.7a)}$$

$$P_{t|t-1} = F P_{t-1|t-1} F' + Q, \quad \text{(C.7b)}$$

where a prime indicates transposition.

Then, the error in the prediction is defined as

$$\eta_{t|t-1} = Y_t - H_t h_{t|t-1}, \quad \text{(C.8)}$$

and its covariance matrix is

$$\chi_{t|t-1} = H_t P_{t|t-1} H_t' - R_t, \quad \text{(C.9)}$$

where R_t is the covariance matrix of the added noise that the approach of Mariano and Murasawa (2003) uses to treat missing values in equation (C.3). Hence, the Gaussian log-likelihood

function can be evaluated as

$$l_t = -\frac{1}{2}\ln(2\pi|\chi_{t|t-1}|) - \frac{1}{2}\eta'_{t|t-1}(\chi_{t|t-1})^{-1}\eta_{t|t-1}. \tag{C.10}$$

The next step in the Kalman filter is updating the estimation with the Kalman gain, typically defined as $K_t = P_{t|t-1}H'_t(\chi_{t|t-1})^{-1}$, such that

$$h_{t|t} = h_{t|t-1} + K_t\eta_{t|t-1}, \tag{C.11}$$

$$P_{t|t-1} = P_{t|t-1} - K_tH_tP_{t|t-1}. \tag{C.12}$$

The initial parameters used to start the filter are typically a vector of zeros for (C.11) and a diagonal matrix for (C.12); the parameters that ultimately minimize the log-likelihood function in equation (C.10) are used as model fit parameters. The minimization is carried out through the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (Liu and Nocedal, 1989). Finally, missing values can be added at the end for doing forecast. This operation can be done since if not observed at period τ, the updating equation will be $h_{\tau|\tau} = h_{\tau|\tau-1}$, which will not change the dynamics of the model.

Comparison models and performance metrics

To quantify the ability of the DFM to predict the dynamics of homicides with firearm, its performance is compared to other benchmark models. Specifically, we examined models from two different families: AI and classic autoregressive models. Details about model implementation are presented in Appendix A and related codes are available at Github.

Within AI models, we considered two tree-based family models (random forest, RF, following Berk et al. (2009), and gradient boosting trees, GBOOST, following Kim et al. (2018)) and a long short-term memory recurrent neural network (LSTM), following Muthamizharasan and Ponnusamy (2022). In all cases, we created predictions by accounting for time variations in the availability of data. Specifically, in each time-segment, the models were trained using all the covariates available and predictions were made using models trained up to the latest available datapoint.

With respect to classical autoregressive models, we considered the univariate ARIMA model as the standard benchmark for time-series modeling. To acknowledge variations in levels and trends of the homicides with firearm series, we also implemented a Holt-Winters model (Winters,1969, Gardner Jr. and McKenzie, 1985). For completeness, we considered a VAR modeling as the benchmark for multivariate analysis, following Parkin et al. (2020) – Results are presented in Appendix B.

The main measurement used for comparison is the mean absolute error (MAE) of the model for the entire year at every month of prediction. To compare the predictive powers of the methods, we considered the HLN-statistic based on the difference in the model residuals, established by Diebold and Mariano (1995) and refined by Harbey et al. (1997) to deal with short series. To assess the ability of the forecasts to accurately anticipate changes in the series' directionality we used the PT-statistic (Pesaran and Timmermann, 1992), by which one can monitor the extent to which a model can anticipate the sign of the variation between two different time-steps of a series.

4.3 — Results

In-sample model estimation

The smallest model that achieved convergence in the optimization of the parameters over the entire sample period consisted of the following four variables: monthly homicides with firearm, provisional quarterly estimates of homicides with firearm, deaths in incidents with guns from Guns Violence Archive, and media output of "homicide" + "shot". To improve the model accuracy, we systematically added one of the remaining variable if: i) its inclusion did not affect convergence, and ii) the new loading factor of the DFM was statistically significant at a confidence level of $\alpha = 0.05$. Such a procedure for model enrichment was conducted on both lagged and non-lagged variable. The final model also included the variables of homicides from three cities and background checks with seven months of lag. The b_i's parameters in equation (C.2) and their p-values are shown in Table 2.

The fitted model explained 59.80% of the variance, as estimated through linear regression of monthly homicides with firearm and the common factor. The resulting common factor, along with the reconstructed series of monthly homicides with firearm, are shown in Figure 4.3 for the sample period.

To ascertain test the robustness and appropriateness of the obtained model, we inspected the features of the residuals of the decomposed series. Specifically, we checked whether the residuals were normally distributed and serially uncorrelated. With respect to nornality, we utilized four different tests settled in the literature: the Kolmogorov-Smirnov test (Massey Jr., 1951) and its corrected version by Lilliefors (1967), the non-parametric entropy-based test by Vasicek (1976), and the Shapiro-Francia test (Shapiro and Francia, 1972; Royston, 1993) for censored data after censoring extreme data. To test for serial correlation, we employed both the Box-Pierce (Box and Pierce, 1970) and the Ljung-Box (Ljung and Box, 1978) tests.

Results in Table 3 offer evidence in favor of normality and non-serial correlation. First, all of the normality tests yield large p-values for the null that the residuals are Gaussian, with the

unique exception of background checks. Such a departure from normality should not be treated as a major concern. In fact, according to Durbin and Koopman (2012), if the true distribution of the error is non-Gaussian, then the Kalman filter would still provide the minimum variance linear unbiased estimator of the state variables, especially when only a handful of the residuals depart from normality (Barigozzi and Luciani, 2019). Second, all the residuals are serially uncorrelated, with the unique exception of those of quarterly annualized estimates of homicides with firearm – which should be expected given the time resolution of this variable. Similar to isolated lack of normality, serial correlation of a few variables have limited effect on the estimators and forecast in DFMs Stock and Watson (2002).

Out-of-sample comparison

An out-of-sample pseudo-real-time exercise was performed to evaluate DFM predictions. For each period of the sample history, a database was created with data available in that period. For each database, we estimated model parameters and recorded the corresponding forecasts. We compared forecasts with the true values and computed the MAE for each forecasted year at every month. The DFM was utilized to forecast future homicides with firearms and to provide missing values in the previous (backcasting) and present years (nowcasting). For example, in March 2017, the last available data for homicides with firearm is for December 2015, although information about other covariates can be available until March 2017. Through the DFM, we backcast from January 2016 until February 2017, nowcast March 2017, and forecast from April 2017 till December 2017. In what follows, we organize our comparisons between DFM predictions and real data in terms of MAE values for the previous (backcasting) and present year (backcasting, nowcasting, and forecasting for all months except January and December).

The same analysis was carried out for AI models. Since these models do not systematically impute missing values, we interpolated quarterly data using splines with the *imputeTS* R package. For the case of RF and GBOOST, multiple models were trained according to the availability of data. Should a variable not be available at a given month, it was excluded from the model training. For the case of LSTM, the procedure was based on training with sequences of three periods, using the *Keras* R package. The parameters of the RF and GBOOST models were identified using cross-validation over the whole sample period; for the latter model we used a Gaussian loss function. The networks' composition of the LSTM was based on training over the whole sample. A dropout layer was included after each LSTM layer, with a final time-distributed layer, and an *adam* optimizer (Kingma and Ba, 2015) was used. In addition to AI models, an ARIMA model was included as a benchmark. The number of parameters was chosen through the Akaike information criterion, yielding an ARIMA(3,0,3) as the model that minimized the score. Since new data for homicides with firearm appears in December of every year, such a univariate model provides the same prediction for any month from January to December.

The MAE associated with predictions of the past and present year at every month is shown in Figure 4.4. With respect to backcasting the series of homicides with firearm in the previous year, the DFM outperformed every other model, which all showed a comparable performance to an ARIMA. Likewise, the DFM performed better than any other model in the prediction of homicides with firearm in the present years. Importantly, the DFM shows an improved learning ability than any other model, whereby its MAE decays faster than any other model as a function of the month within the year. In other words, the DFM is effective in using information about homicides with firearm and all its covariates in the past to draw accurate predictions in the present year. Additionally, we performed a comparison between DFM and alternative autoregressive models (Holt-Winters and VAR models), whose results, displayed in Appendix B, offer further backing to the predictive power of DFM for backcasting, nowcasting, and forecasting homicides with firearm, together with up to five alternative approaches for RF and GBOOST models, displayed in Appendix C and also with results in favor to the DFM.

The qualitative observations drawn from the study of Figure 4.4 were further supported by statistical analyses using the HLN-statistic, as shown in Table 4. We tested whether the accuracy of the DFM was better than any of the other four models (RF, GBOOST, LSTM, and ARIMA). In agreement with our expectations, we registered a consistent improvement in forecasting using DFM, whereby any comparison yielded a significant difference in the HLN-statistic. With regards to backcasting, the DFM offered improved predictive capacity, yet, some of the comparisons failed to reach statistical significance. In particular, comparisons with ARIMA and GBOOST pointed at a marginal improvement bestowed by DFM.

We also tested the ability of the models to capture changes in directionality through the PT-statistic, that is, we studied the extent to which they were able to forecast changes in direction over different prediction periods. In particular, the analysis of Table 5 offers strong support for the systematic ability of DFM to predict the changes in directionality of homicides with firearm. In addition to reliable predictions of changes in directionality for the first backcast period, the DFM also yielded some predictive abilities for longer backcast periods. Importantly, the DFM was the only model that was able to reliably predict such changes for any forecast horizon[4].

4.4 — Conclusions

In the 2010s, violent crimes exhibited a downward trend in the US (Friedman et al., 2017). However, some types of crimes, mainly those related to firearms, have increased in the period 2015-2020 (Albrecht, 2022), thereby attracting media attention and fueling gun policy debates (Barry et al., 2019). The spark in homicides with firearm during the lockdown in 2020 was the dramatic culmination of such a trend, which, however, remained officially undetected for two

[4]All data and codes are available at Github.

years – until official data was released. Such a systematic delay represents a major hurdle for policymakers and practitioners to promptly react to changes in violence. Establishing reliable, statistical tools to predict homicides with firearm is a key, open challenge.

Standard methods, like univariate autoregressive models, are known to yield unreliable predictions, whose error grows over time (Huang et al., 2020). Multivariate vector autoregressive models could, in principle, stabilize error growth, but the limited length of the time-series challenges the estimation of salient model parameters. The short length of the time-series also hampers the use of AI methods, which may overfit the data and fail to capture underlying patterns. Such an issue is further exacerbated by the wide difference in the range of the available time-series, which further restricts the dataset available for training. To address these issues, we propose a dynamic factor model, as a parsimonious monthly representation of the dataset that is updated in real time as any new information about homicides with firearm and any other explanatory variable (provisional quarterly estimates of homicides with firearm, deaths in incidents with guns from Guns Violence Archive, and media output of "homicide" + "shot") becomes available.

The investigation carried out in this effort bears key methodological and practical insights that should be highlighted. From a methodological point of view, we find that decomposing the time-series of homicides with firearm into common and idiosyncratic components through a dynamic factor model yields superior predictions than standard autoregressive and AI models.The superior performance of the dynamic factor model is likely due to asynchronous reporting calendar for official mortality statistics by the CDC, which strain the applicability of the existing machinery. The dynamic factor model thrives on this asynchronicity, where its Kalman filter allows for the incorporation of any information as it becomes available, thereby easing the processes of backcasting, nowcasting, and forecasting.

Despite their increased computational costs, AI techniques failed to contribute significant improvement with respect to a benchmark ARIMA model, in contrast with the dynamic factor model that systematically outperformed ARIMA in backcasting, nowcasting, and forecasting in terms of HLN-statistic (Harvey et al., 2017). The improved ability to detect trends of the dynamic factor model comes with the further additional advantage of reliably anticipating changes in directionality in the time-series of homicides with firearms, a key element to effective policymaking. Through PT-statistic (Pesaran and Timmermann, 1992), we demonstrated an improved ability to detect such changes for both backcasting and nowcasting, in contrast with any other model that can only be reliably used for very narrow backcasting periods.

To our surprise, a new research has been recently conducted by AH firm (AH Datalitycs, 2022), as highlighted by the New York Times (New York Times, 2022), pointing at a reduction of homicides this year. Gun deaths, injuries, and mass shootings are also down this year, compared to the previous year. The New York Times provides an interesting interpretation of the findings, relating to a return to normalcy after a long pandemic, and touches on some

"bad news bias" that have prevented such a funding to reach a wider audience. Data by the CDC regarding this very same time period will not be available until December 2023: our new approach to prediction of homicides with firearm could be key in accelerating the validation stage of any new findings like those by AH firm. More in general, the superior performance of the dynamic factor model has important, practical implications for policymakers and practitioners who are tasked with making timely decisions on pressing topics around violence with limited and often outdated data. In this vein, the model may be used to fill data gaps and anticipate outbreaks of violence, thereby offering a concrete aid to evidence-based interventions.

Our approach is not free of limitations. One of the key shortcomings of the proposed dynamic model is that, as a single index model, it relies on unique common dynamics. In principle, multiple grouped dynamics might coexist when working with large datasets, so that more than one common factor would be needed. In such a case, homicides with firearm could be associated with more than a single common factor, thereby challenging the use of a dynamic factor model with a single factor for reliable predictions. Likewise, the loading factors of the dynamic factor model could be time-varying; for example, during a certain period, homicides with firearm might be closely related to social unrest and in another period, they might be tied to gun-related dynamics. Such time-varying dynamics might have occured in the last section of the observation, entailing the COVID-19 lockdown, when it is tenable that additional social and economic variables might have played a role. In addition to these methodological drawbacks, we should acknowledge limitations in the process of data curation, whereby data from media can be noisy and crowdsource databases are not officially verified, leading to potentially inaccurate estimations.

In the future, several research directions can be pursued. First, Bayesian estimation following Del Negro and Otrok (2008), or extended/unscented Kalman filters can be leveraged to cope with richer dynamics and nonlinear behaviors. Second, the approach can also be adapted to a spatio temporal setting, as in Lopes et al. (2011), working with time-series of homicides with firearm for distinct US states. Lastly, we envision the inclusion of further data, like the Quarterly Tables from NIBRS (FBI, 2022b) to enrich the existing dataset and enhance the accuracy of our predictions.

4.5 — References

AH Datalitycs., 2022. YTD Murder comparison. (https://www.ahdatalytics.com/dashboards/ytd-murder-comparison/)

Albrecht, J. F., 2022. Increasing gun and community violence in the United States: Causes and analyses. In Understanding and preventing community violence (pp. 119–130). Springer.

Alonso Brito, G. R., Rivero Villaverde, A., Lau Quan, A., and Ruíz Pérez, M. E., 2021.

Comparison between SARIMA and Holt–Winters models for forecasting monthly streamflow in the western region of Cuba. *SN Applied Sciences*, 3 (6), 1–12.

Altindag, D. T., 2014. Crime and international tourism. *Journal of Labor Research*, 35, 1-14.

Baker, S. R., Bloom, N., and Davis, S. J., 2016. Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131 (4), 1593–1636.

Barigozzi, M., and Luciani, M., 2019. Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the EM algorithm. arXiv preprint arXiv:1910.03821 .

Barr, R., and Pease, K., 1990. Crime placement, displacement, and deflection. *Crime and Justice*, 12, 277-318.

Barreras, F., Daz, C., Riascos, A., and Riber, M., 2016. Comparison of different crime prediction models in Bogotá. Documentos CEDE , 34 .

Barry, C. L., Stone, E. M., Crifasi, C. K., Vernick, J. S., Webster, D. W., and McGinty, E. E., 2019. Trends in public opinion on US gun laws: majorities of gun owners and non–gun owners support a range of measures. *Health Affairs*, 38 (10), 1727–1734.

Berk, R., 2008. Forecasting methods in crime and justice. *Annual Review of Law and Social Science*, 4 , 219–238.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., and Ahlman, L., 2009. Forecasting murder within a population of probationers and parolees: a high stakes application of statistical learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172 (1), 191–211.

Beshears, J., Choi, J. J., Fuster, A., Laibson, D., and Madrian, B. C., 2013. What goes up must come down? Experimental evidence on intuitive forecasting. *American Economic Review* , 103 (3), 570–74.

Blumstein, A., and Rosenfeld, R., 2008. Factors contributing to US crime trends. In Understanding crime trends: Workshop report (Vol. 2, pp. 13–44). National Academies Press.

Box, G. E., & Pierce, D. A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65 (332), 1509–1526.

Breiman, L., 2001. Random forests. *Machine Learning*, 45 (1), 5–32.

Brockwell, P. J., and Davis, R. A., 2009. Time series: theory and methods. Springer Science and Business Media.

Bureau of Economic Analysis., 2022. GDP personal income table. BEA's National Economic Account. ($https://apps.bea.gov/iTable/index_nipa.cfm$)

Bureau of Justice Statistics., 2017. National crime victimization survey. Inter-university Consortium for Political and Social Research.

Bureau of Justice Statistics., 2022. National Incident-Based Reporting System, 2020: Extract Files. Inter-university Consortium for Political and Social Research.

Carvalho, J., Santos, J., Torres, R., Santarém, F., & Fonseca, C., 2018. Tree-based methods: Concepts, uses and limitations under the framework of resource selection models. *Journal of Environmental Informatics*, 32 (2).

Centers for Disease Control and Prevention., 2021. National Vital Statistics System, Mortality 1999-2020 on CDC WONDER online database. Multiple cause of death files, 1999-2020. (https://wonder.cdc.gov/ucd-icd10.html)

Cesario, E., Catlett, C., and Talia, D., 2016. Forecasting crimes using autoregressive models. In 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Comput-ing, 14th International Conference on Pervasive Intelligence and Computing, 2nd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech) (pp. 795–802). IEEE.

Chatfield, C., and Yar, M., 1988. Holt-Winters forecasting: some practical issues. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 37 (2), 129–140.

Chen, X., Cho, Y., and Jang, S. Y., 2015. Crime prediction using Twitter sentiment and weather. In 2015 Systems and Information Engineering Design Symposium (pp. 63–68). IEEE.

Cheng, C., and Long, W., 2022. The effect of highly publicized police killings on policing: Evidence from large US cities. *Journal of Public Economics*, 206 , 104557.

Cherian, J., and Dawson, M., 2015. RoboCop: Crime classification and prediction in San Francisco. *Forest*, 15 .

Congress.Gov., 2019. Foundations for evidence-based policymaking Act of 2018. (https://www.congress.gov/bill/115th-congress/house-bill/4174)

Cook, P. J., and Ludwig, J., 2006. The social costs of gun ownership. *Journal of Public Economics*, 90 (1-2), 379–391.

Deadman, D., et al., 2001. Forecasting trends in recorded crime. University of Leicester.

DeLisle, J. R., Never, B., and Grissom, T. V., 2020. The big data regime shift in real estate. *Journal of Property Investment and Finance*, 38(4), 363-395.

Del Negro, M., and Otrok, C., 2008. Dynamic factor models with time-varying parameters: measuring changes in international business cycles. FRB of New York Staff Report(326).

Devi, J. V., and Kavitha, K., 2021. Automating time series forecasting on crime data using

RNN-LSTM. *International Journal of Advanced Computer Science and Applications*, 12 (10).

Diebold, F. X., and Mariano, R. S., 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13 (3), 253–263.

Drake, E. K., Aos, S., and Miller, M. G., 2009. Evidence-based public policy options to reduce crime and criminal justice costs: Implications in Washington State. *Victims and Offenders*, 4 (2), 170–196.

Durbin, J., and Koopman, S. J., 2012. Time series analysis by state space methods (Vol. 38). OUP Oxford.

Estrada, M. A. R., and Ndoma, I., 2014. How crime affects economic performance: The case of Guatemala. *Journal of Policy Modeling*, 36(5), 867-882.

FBI., 2019. 2019 Crime in the United States. (https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in -the-u.s.-2019/topic-pages/violent-crime)

FBI., 2022a. NICS firearm checks: month/year. (https://www.fbi.gov/file-repository/nics_firearm _checks_-_month_year.pdf/view)

FBI., 2022b. Quarterly Uniform Crime Report. (https://crime-data-explorer.fr.cloud.gov/pages/ explorer/crime/quarterly) 19

Feng, M., Zheng, J., Han, Y., Ren, J., and Liu, Q., 2018. Big data analytics and mining for crime data analysis, visualization and prediction. In International conference on brain inspired cognitive systems (pp. 605–614). Springer.

Fox, J. A., and Swatt, M. L., 2009. Multiple imputation of the supplementary homicide reports, 1976–2005. *Journal of Quantitative Criminology*, 25 (1), 51–77.

Friedman, M., Grawert, A. C., and Cullen, J., 2017. Crime trends, 1990-2016. Brennan Center for Justice at New York University School of Law New York, NY.

Gamma, A., Schleifer, R., Weinmann, W., Buadze, A., and Liebrenz, M., 2016. Could Google Trends be used to predict methamphetamine-related crime? An analysis of search volume data in Switzerland, Germany, and Austria. *PLoS ONE* , 11 (11).

Gardner Jr, E. S., and McKenzie, E., 1985. Forecasting trends in time series. *Management Science*, 31 (10), 1237–1246.

Giffords Law Center., 2022. Surging in Gun Violence. (https://giffords.org/lawcenter/report/surging -gun-violence-where-we-are-how-we-got-here-and-where-we-go/)

Gorr, W., Olligschlaeger, A., and Thompson, Y., 2003. Short-term forecasting of crime. *International Journal of Forecasting*, 19 (4), 579–594.

Gramlich, J., 2022. What the data says about gun deaths in the US. Pew Research Center .

Guerron-Quintana, P. A., Khazanov, A., & Zhong, M., 2021. Nonlinear Dynamic Factor Models. Unpublished manuscript.

Guns Violence Archive., 2022. Guns Violence Archive database.
(https://www.gunviolencearchive.org/query)

Harvey, A. C., 1990. Forecasting, structural time series models and the Kalman filter. Cambridge University Press.

Harvey, D., Leybourne, S., and Newbold, P., 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13 (2), 281–291.

Hochreiter, S., 1991. Untersuchungen zu dynamischen neuronalen netzen. Diploma, Technische Universität München, 91 (1).

Hoffman, C. Y., Hinkle, J. C., and Ledford, L. S., 2021. Beyond the "Ferguson Effect" on crime: Examining its influence on law enforcement personnel. CJC Publications. 36 .

Holt, C. C., 1957. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20 (1), 5–13.

Huang, Y., Xu, C., Ji, M., Xiang, W., and He, D., 2020. Medical service demand forecasting using a hybrid model based on arima and self-adaptive filtering method. *BMC Medical Informatics and Decision Making*, 20 (1), 1–14.

Ipsos., 2022. FiveThirtyEight/Ipsos 2022 Election Tracking Survey. (https://www.ipsos.com/en-us/ news-polls/FiveThirtyEight-2022-midterm-election)

Kim, S., Joshi, P., Kalsi, P. S., and Taheri, P., 2018. Crime analysis through machine learning. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 415–420).

Kingma, D. P., and Ba, J., 2015. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations. ICLR.

Lang, M., 2013. Firearm background checks and suicide. *The Economic Journal* , 123 (573), 1085– 1099.

Lilliefors, H. W., 1967. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62 (318), 399–402.

Liu, D. C., and Nocedal, J., 1989. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45 (1), 503–528.

Ljung, G. M., and Box, G. E., 1978. On a measure of lack of fit in time series models. *Biometrika*, 65 (2), 297–303.

Lopes, H. F., Gamerman, D., and Salazar, E., 2011. Generalized spatial dynamic factor models.

Computational Statistics and Data Analysis, 55 (3), 1319–1330.

Mancik, A. M., Hawk, S. R., Jarvis, J. P., and Regoeczi, W. C., 2021. Evaluating fluctuations in homicide: Crowdsourcing trends and assessing sentiments of change. *Journal of Crime and Justice*, 44 (5), 553–578.

Mariano, R. S., and Murasawa, Y., 2003. A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18 (4), 427–443.

Massey Jr, F. J., 1951. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46 (253), 68–78.

McDowall, D., 2002. Tests of nonlinear dynamics in US homicide time series, and their implications. *Criminology*, 40 (3), 711–736.

Meskela, T. E., Afework, Y. K., Ayele, N. A., Teferi, M. W., and Mengist, T. B., 2020. Designing time series crime prediction model using long short-term memory recurrent neural network. *International Journal of Recent Technology and Engineering*, 9(4), 402-405.

Mitchell, T. M., and Mitchell, T. M., 1997. Machine learning (Vol. 1) (No. 9). McGraw-hill New York.

Morgan, R. E., and Thompson, A., 2021. Criminal victimization, 2020 (Vol. 4). Bureau of Justice Statistics.

Muthamizharasan, M., and Ponnusamy, R., 2022. Forecasting Crime Event Rate with a CNN-LSTM Model. In Innovative data communication technologies and application (pp. 461–470). Springer.

Nasridinov, A., Ihm, S.-Y., and Park, Y.-H., 2013. A decision tree-based classification model for crime prediction. In Information technology convergence (pp. 531–538). Springer.

Nau, R., 2014. The mathematical structure of ARIMA models. Duke University Online Article, 1 (1), 1–8.

Neumayer, E., 2003. Good policy can lower violent crime: Evidence from a cross-national panel of homicide rates, 1980–97. *Journal of Peace Research*, 40 (6), 619–640.

New York Times., 2022. A Continuing Drop in Murders. (https://www.nytimes.com/2022/12/30/briefing/crime-murders-us-decline.html)

OECD., 2020. How can governments leverage policy evaluation to improve evidence informed policy making ? (https://www.oecd.org/gov/policy-evaluation-comparative-study-highlights.pdf)

Parker, K. F., Mancik, A., and Stansfield, R., 2017. American crime drops: Investigating the breaks, dips and drops in temporal homicide. *Social Science Research*, 64 , 154–170.

Parkin, W. S., Bejan, V., and Hickman, M. J., 2020. Police, public and community violence:

Exploring the relationships between use of deadly force, law enforcement killed, and homicide rates in the United States. *Criminology, Criminal Justice Law and Society*, 21 , 1.

Pesaran, M. H., and Timmermann, A., 1992. A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics*, 10 (4), 461–465.

Phillips, D. P., and Hensley, J. E., 1984. When violence is rewarded or punished: The impact of mass media stories on homicide. *Journal of Communication*, 10 (3), 101–116.

Phillips, M. D., 2016. Time series applications to intelligence analysis: a case study of homicides in Mexico. *Intelligence and National Security*, 31 (5), 729–745.

Police Executive Research Forum., 2021. Survey on 2020 gun crime and firearm recoveries. (https://www.policeforum.org/criticalissues26jan21)

Pratt, T. C., and Lowenkamp, C. T., 2002. Conflict theory, economic conditions, and homicide: A time-series analysis. *Homicide Studies*, 6 (1), 61–83.

Regoeczi, W. C., Banks, D., Planty, M., Langton, L., and Warner, M., 2014. The nation's two measures of homicide. Sociology and Criminology Faculty Publications. 124 .

Roman, J. K., and Cook, P., 2021. Improving data infrastructure to reduce firearms violence. Chicago, IL: NORC at the University of Chicago.

Rosenfeld, R., and Fornango, R., 2007. The impact of economic conditions on robbery and property crime: The role of consumer sentiment. *Criminology*, 45 (4), 735–769.

Rosenfeld, R., and Fox, J. A., 2019. Anatomy of the homicide rise. *Homicide Studies*, 23 (3), 202–224.

Rosenfeld, R., and Lopez, E., 2020. Pandemic, social unrest, and crime in US cities. Council on Criminal Justice.

Royston, P., 1993. A pocket-calculator algorithm for the shapiro-francia test for non-normality: An application to medicine. *Statistics in Medicine*, 12 (2), 181–184.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1986. Learning representations by back-propagating errors. *Nature*, 323 (6088), 533–536.

Sargent, T. J., Sims, C. A., et al., 1977. Business cycle modeling without pretending to have too much a priori economic theory. In New methods in business cycle research (Vol. 1, pp. 145–168). Minneapolis: Federal Reserve Bank of Minneapolis.

Shapiro, S. S., and Francia, R., 1972. An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67 (337), 215–216.

Shoesmith, G. L., 2013. Space–time autoregressive models and forecasting national, regional and state crime rates. *International Journal of Forecasting*, 29 (1), 191–201.

Sims, C. A., 1980. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1–48.

Soybilgen, B., 2020. Identifying US business cycle regimes using dynamic factors and neural network models. *Journal of Forecasting*, 39 (5), 827–840.

Stock, J. H., 1994. Unit roots, structural breaks and trends. *Handbook of Econometrics*, 4 , 2739–2841.

Stock, J. H., and Watson, M. W., 1991. A probability model of the coincident economic indicators. In Leading economic indicators: New approaches and forecasting records. Cambridge University Press.

Stock, J. H., and Watson, M. W., 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97 (460), 1167–1179.

Strom, K. J., and Smith, E. L., 2017. The future of crime data: The case for the National Incident-Based Reporting System (NIBRS) as a primary data source for policy evaluation and crime analysis. *Criminology and Public Policy*, 16 (4), 1027–1048.

Vasicek, O., 1976. A test for normality based on sample entropy. *Journal of the Royal Statistical Society: Series B (Methodological)*, 38 (1), 54–59.

Vezhnevets, A., and Barinova, O., 2007. Avoiding boosting overfitting by removing confusing samples. In European Conference on Machine Learning (pp. 430–441).

Wallace, L. N., 2015. Responding to violence with guns: Mass shootings and gun acquisition. *The Social Science Journal* , 52 (2), 156–167.

Wallis, K. F., 1986. Forecasting with an econometric model: The 'ragged edge'problem. *Journal of Forecasting*, 5 (1), 1–13.

Wheeler, A. P., and Kovandzic, T. V., 2018. Monitoring volatile homicide trends across US cities. *Homicide Studies*, 22 (2), 119–144.

Winters, P. R., 1960. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6 (3), 324–342.

Yadav, R., and Sheoran, S. K., 2018. Crime prediction using auto regression techniques for time series data. In 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE) (pp. 1–5). IEEE.

Yim, H.-N., Riddell, J. R., and Wheeler, A. P., 2020. Is the recent increase in national homicide abnormal? Testing the application of fan charts in monitoring national homicide trends over time. *Journal of Criminal Justice*, 66 .

Ying, X., 2019. An overview of overfitting and its solutions. *Journal of Physics: Conference*

Series, 1168 (2).

4.6 — Appendix A

Description of methods used for comparison

AI and ARIMA models

Here, we present a brief description of the AI and ARIMA models that are used in the main manuscript to assess the performance of the DFM. First, we offer some intuition behind tree-based methods (random forest, RF and gradient boosting trees, GBOOST) and then turn our attention to long short-term memory neural networks (LSTM) and the ARIMA model.

RF. Tree-based methods (Breiman, 2001) involve the segmentation/partition of the predictor space into a finite set of regions by minimizing a given loss function. Toward improved robustness, RF and GBOOST produce multiple trees that are combined to yield a single consensus prediction. More specifically, RF builds multiples decision trees on bootstrapped training samples and averages them to obtain the prediction. To avoid the creation of correlated trees that would grow the variance of the final predictor, the algorithm considers a random sample of m predictors among the full available set at each split of any tree. From cross-validation of the whole sample, we set $m = 48$, the number of bootstrapped trees equal to 200, and the minimum node size equal to 5.

GBOOST. GBOOST does not involve bootstrap sampling. Instead, the trees are grown sequentially and each of them is fit to a modified version of the original dataset. Specifically, the algorithm has three parameters: the number of trees B, the shrinkage parameter λ, and the number of splits d in each tree, which controls the complexity of the boosted ensemble. The algorithm implements the following steps:

1. Initialize the algorithm with a null tree ($\hat{f} = 0$) and residuals equal to the observations of the dependent variable ($r_i = y_i$, for all i in the training set).

2. For $b = 1, 2 \ldots, B$, create multiple trees by

   (a) fitting a tree $\hat{f}^b$ with d splits;

   (b) update the tree by adding a shrunken version of the new tree,

$$\hat{f} \leftarrow \hat{f} + \lambda \hat{f}^b,$$

(c) update the residuals

$$r_i \leftarrow r_i - \lambda \hat{f}^b(x_i),$$

where x_i is the value of the independent variable in the training set.

3. Output the boosted model

$$\hat{f} = \sum_{b=1}^{B} \lambda \hat{f}^b.$$

After cross validation with the whole sample, we choose $B = 20$, $\lambda = 0.5$, and $d = 3$.

LSTM. The long short-term memory networks, LSTM, are a particular case of recurrent neural networks (RNN) that were proposed by Rumelhart et al. (1986) to process variable length sequences of inputs. The LSTM network is created with different layers where a back-propagation algorithm (typically through the stochastic gradient descent procedure) is used to estimate the network parameters during the training process. In contrast with standard RNN, the LSTM networks do not suffer from the so-called "long short-term memory problem", as pointed out in Hochreiter (1991), from which during back-propagation the gradient might vanish or explode and therefore artificially affecting the training process. In our case, the *adam* optimizer was used, which is an extension of the standard stochastic gradient descent procedure, to minimize the mean absolute error (MAE). Moreover, the networks composition were first trained with the whole sample, minimizing the error the configuration with two LSTM layers, with 50 and 20 neurons respectively. A dropout layer was included after each LSTM layer, with a dropout rate of 0.5 to avoid overfitting, and a final time-distributed layer was added at the end. The training was performed through 10 epochs.

ARIMA. The univariate autoregressive model assumes that the dynamics of a time series, y_t, is driven by its own past. In particular, assuming the number of lag-observations to be equal to p and mean value to be equal to μ, the time-series can be written as

$$y_t = \mu + a(L)(y_{t-1} - \mu) + \varepsilon_t, \tag{C.13}$$

where L is the lag-operator, $a(L) = (a_1 + a_2 L + ... + a_p L^{p-1})$, and the errors ε_t are i.i.d. following $\mathcal{N}(0, \sigma^2)$. By choosing the order of the moving average to be equal to q and the degree of differencing to be equal to d, we obtain ARIMA(p, d, q), defined as

$$(1 - a_1 - a_2 L - ... - a_p L^{p-1})((1 - L)^d y_t - \mu) = (1 + \theta_1 + \theta_2 L + ... + \theta_q L^q)\varepsilon_t, \tag{C.14}$$

For the time-series of homicides with firearm, we used an ARIMA(3,0,3) as a parsimonious, yet descriptive model; the estimated coefficients and their respective p-values are presented in Table A1.

Table A1. Estimated coefficients for the ARIMA(3,0,3) for homicides with firearm.

	a_1	a_2	a_3	θ_1	θ_2	θ_3
Value	-0.09	0.12	0.95	0.68	0.64	-0.35
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$

Notes. The intercept was not statistically significant and the resulting AIC was −7510.

Alternative autoregressive models

In addition to the ARIMA examined in the main manuscript, we considered two alternative autoregressive models: the Holt-Winters' method, initially proposed by Holt (1957) and extended to account for seasonality by Winters (1960), and the vector autoregressive (VAR) model, the natural extension of univariate models to multivariate settings Sims (1980).

Holt-Winters. The Holt-Winters' method uses exponential smoothing to encode the values of a time-series, y_t, as a combination of three components: the level (l_t), trend (b_t), and seasonal factor (s_t). The three components are determined using smoothing methods as follows:

$$\begin{aligned} l_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}, \end{aligned} \tag{C.15}$$

where α, β, and γ are the corresponding smoothing parameters, and m denotes the frequency of the seasonality. The parameters of the model can be estimated by minimizing the residual sum of squares (in our case, $\alpha = 0.3, \beta = 0.1$, and $\gamma = 0.1$), and l_t, b_t, and s_t can be obtained by simply initializing at the first time-step. For forecasting h time-steps in the future, we implement the following:

$$y_{t+h} = l_t + hb_t + s_{t+h-m(k+1)}, \tag{C.16}$$

where k is the integer part of $(h-1)/m$.

VAR. Let Y_t be a vector of n stationary time- series $y_{1,t}, ..., y_{n,t}$ with expected value ν. The reduced form of a VAR model with p lags, VAR(p), is

$$Y_t = \nu + C(L)(Y_{t-1} - \nu) + u_t, \tag{C.17}$$

where $C(L) = (C_1 + C_2L + ... + C_pL^{p-1})$ is the matrix of lag polynomials and the errors are serially uncorrelated with zero mean and covariance matrix Ω. In our database, the in-sample estima-

tion is performed with only four variables: homicides with firearm, media output on "homicide" + "shot", homicides from three cities, and background checks, while deaths in incidents with guns from the Guns Violence Archive and provisional quarterly annualized estimate rate for homicides with firearm from the CDC are omitted because they start very late The model can be estimated by ordinary least squares in each equation and the lag-length is typically chosen by using model selection criteria (in our case, we used AIC and obtained $p = 3$). In-sample parameter estimates are shown in Table A2.

Table A2. Estimated coefficients for the VAR(3) model.

	HF_1	MOH_1	$H3_1$	BCs_1	HF_2	MOH_2	$H3_2$	BCs_2	HF_3	MO_3	$H3_3$	BCs_3	ν
HF	$4{\cdot}10^{-1}$	$2{\cdot}10^{-4}$	$2{\cdot}10^{2}$	$2{\cdot}10^{-8}$	$3{\cdot}10^{-1}$	$2{\cdot}10^{-5}$	$-7{\cdot}10^{2}$	$-1{\cdot}10^{-9}$	$1{\cdot}10^{-1}$	$1{\cdot}10^{-4}$	$1{\cdot}10^{2}$	$1{\cdot}10^{-2}$	$6{\cdot}10^{-10}$
MOH	$2{\cdot}10^{1}$	$3{\cdot}10^{-1}$	$1{\cdot}10^{5}$	$-8{\cdot}10^{-6}$	$1{\cdot}10^{-2}$	$9{\cdot}10^{-2}$	$-3{\cdot}10^{5}$	$7{\cdot}10^{-6}$	$4{\cdot}10^{-1}$	$3{\cdot}10^{-3}$	$3{\cdot}10^{5}$	$-3{\cdot}10^{-7}$	$-3{\cdot}10^{1}$
H3	$2{\cdot}10^{-6}$	$9{\cdot}10^{-9}$	$2{\cdot}10^{-1}$	$2{\cdot}10^{-12}$	$1{\cdot}10^{-5}$	$-2{\cdot}10^{-9}$	$2{\cdot}10^{-1}$	$-2{\cdot}10^{-12}$	$-1{\cdot}10^{-5}$	$2{\cdot}10^{-8}$	$3{\cdot}10^{-1}$	$7{\cdot}10^{-13}$	$-2{\cdot}10^{-6}$
BCs	$1{\cdot}10^{6}$	$2{\cdot}10^{3}$	$6{\cdot}10^{9}$	$8{\cdot}10^{-1}$	$-1{\cdot}10^{6}$	$4{\cdot}10^{2}$	$-6{\cdot}10^{9}$	$-2{\cdot}10^{-1}$	$1{\cdot}10^{6}$	$2{\cdot}10^{3}$	$-1{\cdot}10^{10}$	$2{\cdot}10^{-1}$	$-4{\cdot}10^{-6}$

Notes. HF refers to monthly homicides with firearm from CDC, H3 to monthly homicides in the three cities, MOH to monthly media output with "homicide"+"shot", and BCs to monthly background checks. Subscript refers to the lag of the variable in the model.

4.7 — Appendix B

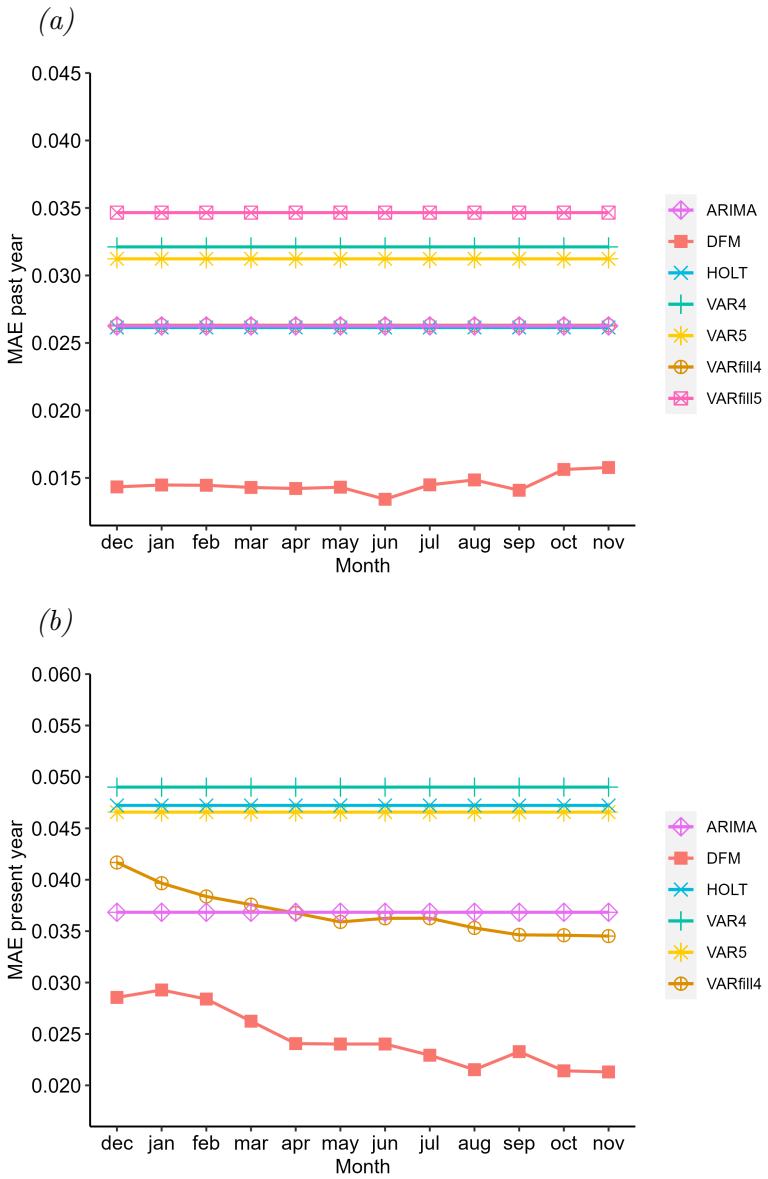
Comparison with alternative autoregressive models

Here, we compare the DFM with Holt-Winters' method and VAR. Similar to ARIMA, the Holt-Winters' method yields the same forecast for every estimation period from January to December as new data for homicides with firearm appear in December every year. With the VAR model, we implemented two different approaches for prediction. In the first, we performed an iterative forecast from the last time period where all observations for all the four variables were available and we refer to these results as VAR4. For predictions after 2017, there is the possibility to include deaths in incidents with guns from the Guns Violence Archive; the VAR model accounting for this extra variable is referred to as VAR5. In the second approach, which we refer to as VAR "fill", we performed a one-step-ahead forecast from the period where all observations were available. For the next period, if some variables became available, we disregarded the forecast of those variables and used observations in their place. In this case, the forecasts computed with the four original variables are called VARfill4, while those that include also deaths in incidents with guns from the Guns Violence Archive are referred to as VARfill5.

The results of the out-of-sample analysis are displayed in Figure B1. As in the case of AI models, the DFM was the only model with better performance than the ARIMA benchmark model. Holt-Winters and VARfill4 performed similarly to the ARIMA model in the previous

years, while just VARfill4 did it in the present years. Finally, VAR4, VAR5, and VARfill5 predictions perform poorly.

Figure B1. Out-of-sample pseudo real-time comparison between DFM and other autoregressive models in the period from January 2008 to December 2020, in terms of MAE: (a) previous and (b) present years.



Notes. Each point represents an MAE value for either the entire year preceding the month at which the prediction is made (a) or the entire year when the prediction is made (b). DFM (red, ■), VAR with four variables and "fill" schema (orange, ⊕), VAR with four variables (turquoise, +), VAR with five variables and "fill" schema (pink, ⊠), VAR with five variables (yellow, ∗), Holt-Winters' method (light blue, ×), and ARIMA(3,0,3) (violet, ⊕). In (b), VAR predictions with five variables with "fill" scheme are not shown for their higher range. VAR and VAR models with "fill" scheme do not contain predictions for years 2013 and 2008, respectively, due to their outlier performance.

4.8 — Appendix C

Accuracy comparison within different approaches for RF and GBOOST models

Due to the fact that RF and GBOOST don't take into account missing values in a consolidated framework as DFM does, several estimation and prediction schemes can be adopted. To ensure the robustness of the results depicted in the main text of this chapter, we carried out the same out of sample exercise in pseudo real time to backcast and nowcast the homicides with firearm for the period between 2008 and 2020 with up to five additional approaches, three for the RF and two for the GBOOST procedure.

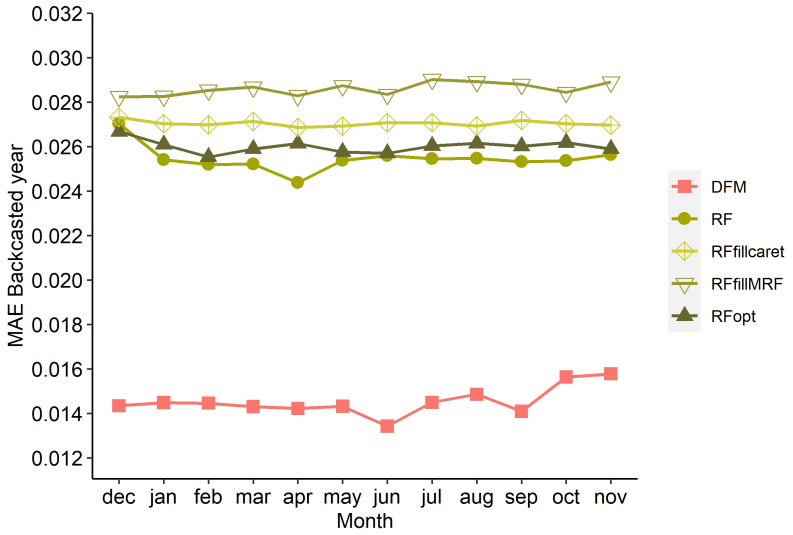
Specifically, within the RF modeling, whereas in the standard approach used in the study the parameters were optimized with the whole sample and kept fixed in the out of sample exercise, an alternative where the optimization was done at every period of estimation. However, these procedures are based on a univariate output, since typically RF trains several covariates to regress (or classify) a single variable as the output. Therefore, two alternatives were also adopted with a multioutput framework. For both, the procedure to create the prediction is different than the one described in the main text. At each estimation period t where homicides with firearm is not available, a prediction for each variable in the next period is made. Thus, a prediction of every variable in the dataset is obtained for $t+1$. If any of the variables is available at $t+1$ its prediction is discarded, so that for $t+2$ the prediction is created with the available data and the forecasted values of the nonavailable data. This procedure is repeated for the whole backcast and nowcast estimations. We refer to this repeated procedure as "fill" scheme. It must be noted, that every time a new variable appeared, the training has to be restarted. In particular, it had to be done at three moments, when the variable of tweets containing the word "homicide" appears, and also with deaths in incidents involving guns and with the appearance of the provisional quarterly estimate of homicide with firearm.

In the first alternative adopted, the Python package *caret* allows being estimated within the multioutput regressor class, which turns into a univariate output estimation for each variable in the data. Regarding the second alternative, Rahman et al. (2017) adapted in the $MultivariateRandomForest$ package in R the joint estimation proposal in Segal and Xiao (2011) where the split function is extended by replacing the node impurity measure with a covariance weighted analog. Nevertheless, to apply thus approach in out case, only the variables recommended by the DFM were used due to the size of the dataset, as the use of more variables did not allow for the required estimation of the inverse of the covariance matrix of the output response matrix. The MAE when predicting the backcasted and nowcasted year every month is shown in Figures C2 and C4. For the backcast exercise, the DFM had a significantly better

performance than any of the RF models. Moreover, the results were similar for the four different approaches with the RF, slightly better that of the univariate response which optimized the parameters just with the whole sample. For the nowcast exercise, a similar conclusion can be deducted but for the first four month where the multipoutput approach with the *caret* package had a slightly better performance although still far from the DFM results.

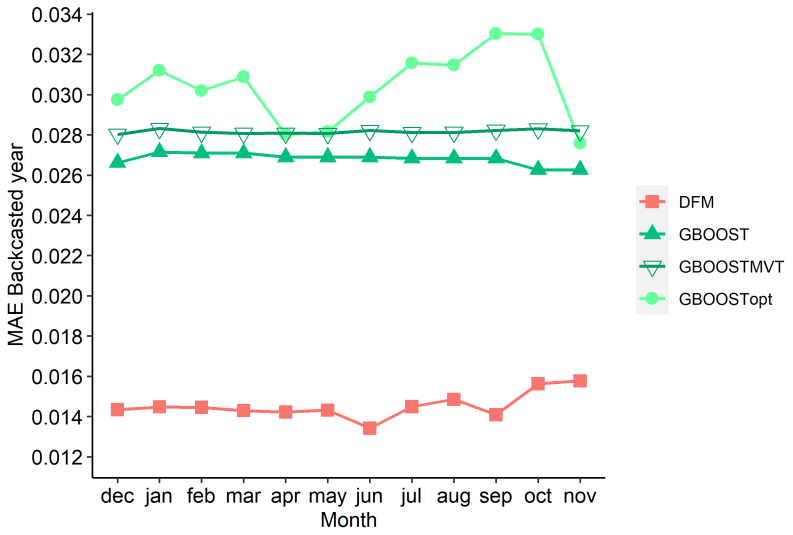
Within the GBOOST approach, the first alternative consisted also of the optimization of the parameters at every estimation and prediction period. On the other hand, in the second alternative procedure adopted the estimation was performed by using the *mvtboost* package in R Miller et al., 2016, an adaptation of the *gbm* package that allows for multioutput response estimation by a different univariate output estimation for each variable, a similar approach to that previously described for the RF when using the *caret* package in Python within the multioutput response class. Figures C1 and C3 represent the MAE when predicting the backcasted and nowcasted year every month. For the backcast exercise, the DFM again had a significantly better performance than any of the GBOOST approaches. The results for the univariate response model that optimized the parameters just with the whole sample where similar to the ones obtained with the multioutput response approach. Furthermore, the univariate output model that optimized the parameters at every period was slightly worse. Regarding the nowcast exercise, again the univariate output model with a single optimization had a better perfomance but for the first months of estimation, where the multioutput one improves its performance. Nevertheless, the results are still far from those from DFM.

Figure C1. Out of sample pseudo real-time 2008-20.



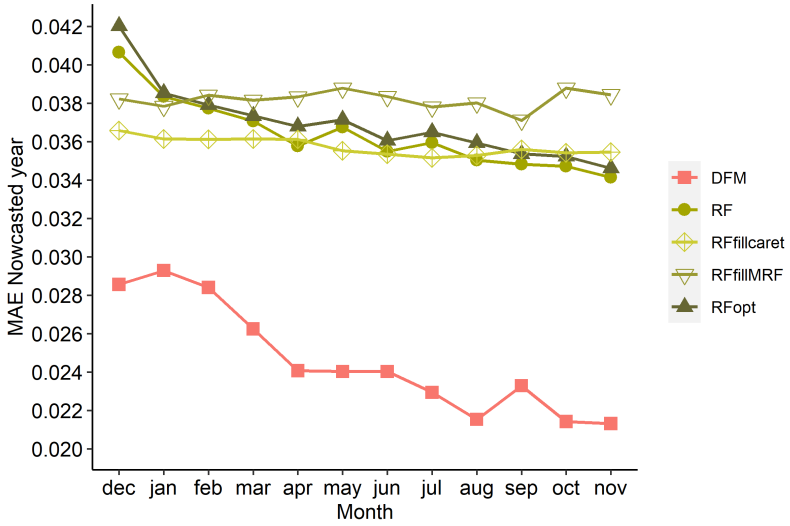
Notes. A value represents the MAE produced for the whole backcasted years at a certain month. MAE series of DFM (red, ■), RF (dark green, •), RF with multioutput from *caret* and "fill" schema (mustard, ⊕), RF with multioutput and "fill" schema (olive green, ▽) and RF with univariate output and periodic optimization (black, ▲).

Figure C2. Out of sample pseudo real-time 2008-20.



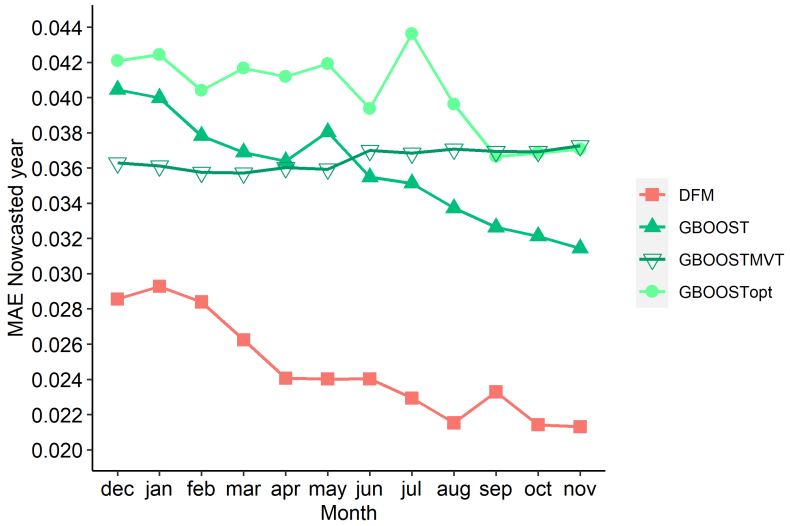
Notes. A value represents the MAE produced for the whole backcasted years at a certain month. MAE series of DFM (red, ■), GBOOST (light green, ▲), GBOOST with multioutput response (pine green, ▽), GBOOST with univariate output and periodic optimization (mint green, •).

Figure C3. Out of sample pseudo real-time 2008-20.



Notes. A value represents the MAE produced for the whole nowcasted years at a certain month. MAE series of DFM (red, ■), RF (dark green, •), RF with multioutput from *caret* and "fill" schema (mustard, ⊕), RF with multioutput and "fill" schema (olive green, ▽) and RF with univariate output and periodic optimization (black, ▲).

Figure C4. Out of sample pseudo real-time 2008-20.



Notes. A value represents the MAE produced for the whole nowcasted years at a certain month. MAE series of DFM (red, ■), GBOOST (light green, ▲), GBOOST with multioutput response (pine green, ▽), GBOOST with univariate output and periodic optimization (mint green, •).

4.9 — Tables

Table 1. Data features

Variable	Frequency	Sample	Delay
Homicides with firearm	monthly	1999M1-2020M12	12-23 months
Provisional estimate of homicides with firearm	quarterly	2017Q1-2020Q4	10 months
Deaths in incidents with guns	monthly	2014M1-2020M12	no delay
Monthly homicides (averaged in three cities)	monthly	2006M1-2020M12	no delay
Google Trends on "homicide"	monthly	2004M1-2020M12	no delay
Background checks	monthly	1999M1-2020M12	no delay
Media output on "homicide"+"shot"	monthly	1999M1-2020M12	no delay
Media output on "riots"	monthly	1999M1-2020M12	no delay
Media output on "unemployment"	monthly	1999M1-2020M12	no delay
Economical Policy Uncertainy index	monthly	1999M1-2020M12	no delay
Google Trends on "gun"	monthly	2004M1-2020M12	no delay
Number of tweets about "homicide"	monthly	2010M1-2020M12	no delay

Notes.Sampling frequency, time-interval, and delays (in months) for each of the variables forming the dataset used to model and predict homicides with firearm at a monthly resolution.

Table 2. Loading factors for the model fitted over the entire sample period from January 1999 to December 2020.

	HF	PQE HF	GVA	H3	MOH	BCs(-7)
b_i	0.34	0.07	0.20	0.21	0.16	0.10
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$

Notes. HF refers to monthly homicides with firearm from CDC, PQE HF to provisional quarterly annualized estimate of homicides with firearm from CDC, GVA to deaths in incidents with guns from Guns Violence Archive, H3 to monthly homicides in the three cities, MOH to monthly media output with "homicide"+"shot", and BCs(-7) to monthly background checks lagged seven periods.

Table 3. Diagnostic tests on the DFM residuals.

	Normality				Serial correlation	
	K-S	L	V	S-F	B-P	L-B
ϵ_t^f	0.036 (0.883)	0.036 (0.554)	0.098 (0.062)	1.315 (0.094)	1.327 (0.249)	1.343 (0.247)
ϵ_{1t}^u	0.077 (0.950)	0.077 (0.775)	0.153 (0.911)	-0.527 (0.701)	9.536 (0.002)	10.252 (0.001)
ϵ_{2t}^u	0.054 (0.439)	0.054 (0.066)	0.091 (0.174)	-0.068 (0.527)	1.274 (0.259)	1.288 (0.256)
ϵ_{3t}^u	0.063 (0.884)	0.063 (0.592)	0.130 (0.445)	-0.543 (0.707)	0.124 (0.725)	0.128 (0.720)
ϵ_{4t}^u	0.043 (0.749)	0.043 (0.320)	0.120 (0.078)	1.720 (0.044)	0.106 (0.745)	0.107 (0.744)
ϵ_{5t}^u	0.042 (0.913)	0.042 (0.624)	0.104 (0.504)	-0.592 (0.723)	2.194 (0.139)	2.231 (0.135)
ϵ_{6t}^u	0.154 (¡0.001)	0.154 (¡0.001)	0.414 (¡0.001)	2.116 (0.017)	2.231 (0.345)	0.135 (0.342)

Notes. ϵ_t^f refers to the error of the common factor and ϵ_{it}^u to the errors of the idiosyncratic components: i=1 for provisional quarterly annualized estimate of homicides with firearm from CDC; i=2 for monthly homicides with firearm from CDC; i=3 for deaths in incidents with guns from Guns Violence Archive; i=4 for monthly media output with "homicide"+"shot"; i=5 for monthly homicides in the three cities; and i=6 for monthly background checks lagged seven periods. Numbers in parentheses are p-values. We use the following acronyms: K-S (Kolmogorov-Smirnov test), L (corrected version of the Kolmogorov-Smirnov by Lilliefors test), V (non-parametric entropy-based test by Vasicek), S-F (Shapiro-Francia test), B-P (Box-Pierce test), and L-B (Ljung-Box test).

Table 4. Statistical analysis of DFM performance in backcasting and forecasting against other AI models and ARIMA using the HLN-statistic.

	DFM vs RF	DFM vs GBOOST	DFM vs LSTM	DFM vs ARIMA
		Backcast		
		$h = 1$		
HLN	**-3.01**	-1.21	**-3.00**	-1.54
p-value	$< 10^{-2}$	0.23	$< 10^{-2}$	0.12
		$h = 2$		
HLN	**-2.23**	-1.42	**-3.64**	**-1.54**
p-value	0.03	0.15	$< 10^{-2}$	0.01
		$h = 3$		
HLN	-1.77	-1.92	-1.12	-1.82
p-value	0.08	0.05	0.07	0.07
		Forecast		
		$h = 1$		
HLN	**-4.94**	**-5.54**	**-5.32**	**-5.12**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
		$h = 2$		
HLN	**-4.78**	**-4.87**	**-4.57**	**-4.55**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
		$h = 3$		
HLN	**-4.78**	**-4.80**	**-4.74**	**-4.70**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
		$h = 4$		
HLN	**-4.95**	**-5.89**	**-4.51**	**-4.22**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
		$h = 5$		
HLN	**-5.62**	**-5.59**	**-4.56**	**-4.17**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
		$h = 6$		
HLN	**-5.01**	**-5.33**	**-4.43**	**-4.04**
p-value	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$

Notes. The comparisons are carried out for different forecast/backcast horizons (h); bold values indicate a significant statistic at $\alpha = 0.05$.

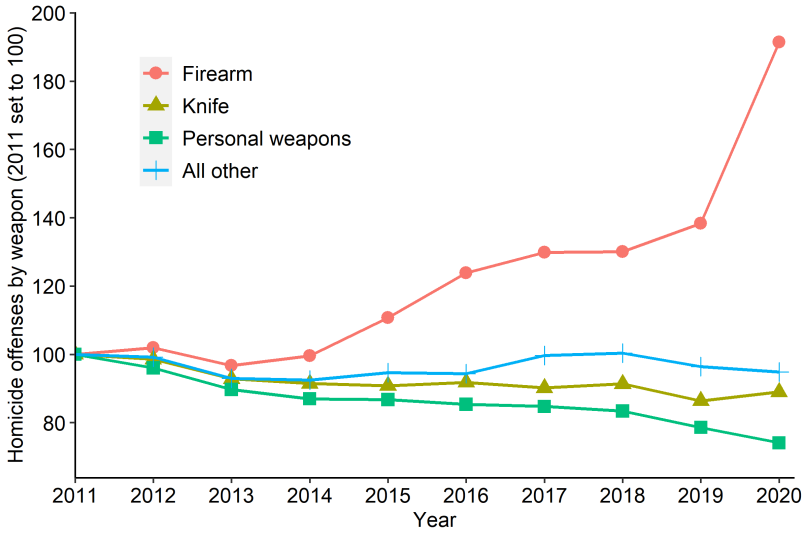
Table 5. Statistical analysis of direction predictive accuracy of DFM, AI models, and ARIMA using the PT-statistic.

	DFM	RF	GBOOST	LSTM	ARIMA
			Backcast		
			h =1		
PT	**2.38**	**1.83**	1.14	**3.06**	**1.83**
p-value	$< 10^{-2}$	0.03	0.13	$< 10^{-2}$	0.03
			h = 2		
PT	1.06	0.61	0.61	0.10	-0.61
p-value	0.14	0.27	0.27	0.46	0.73
			h = 3		
PT	**2.34**	0.61	**1.83**	-0.61	-1.83
p-value	$< 10^{-2}$	0.27	0.03	0.73	0.97
			Forecast		
			h = 1		
PT	**5.42**	0.72	1.02	-2.32	0.00
p-value	$< 10^{-2}$	0.24	0.15	0.99	0.5
			h = 2		
PT	**4.67**	-1.02	-1.41	-0.61	-1.99
p-value	$< 10^{-2}$	0.85	0.92	0.73	0.98
			h = 3		
PT	**4.34**	0.20	0.51	-0.71	-2.27
p-value	$< 10^{-2}$	0.42	0.31	0.76	0.99
			h = 4		
PT	**3.82**	0.28	0.58	-2.23	-1.64
p-value	$< 10^{-2}$	0.39	0.28	0.99	0.95
			h = 5		
PT	**3.90**	0.94	0.89	-0.75	-1.85
p-value	$< 10^{-2}$	0.17	0.19	0.77	0.97
			h = 6		
PT	**3.85**	0.27	-0.10	-0.68	-1.18
p-value	$< 10^{-2}$	0.39	0.54	0.75	0.88

Notes. The comparisons are carried out for different forecast/backcast horizons (h); bold values indicate a significant statistic at $\alpha = 0.05$.

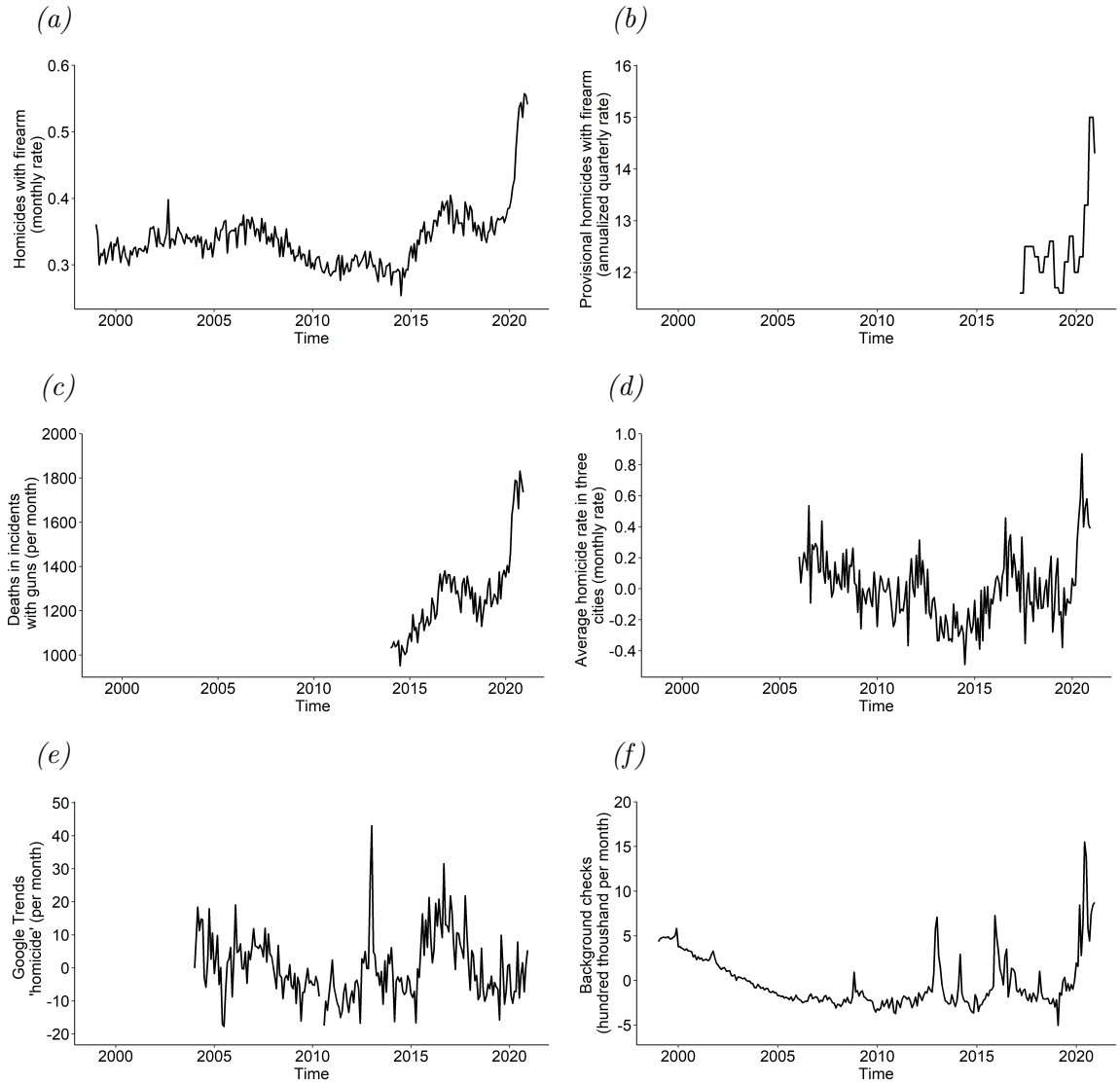
4.10 — Figures

Figure 4.1. Relative changes in homicide offenses rates by type of weapon in the period from January 2011 to December 2020, from NIBRS.



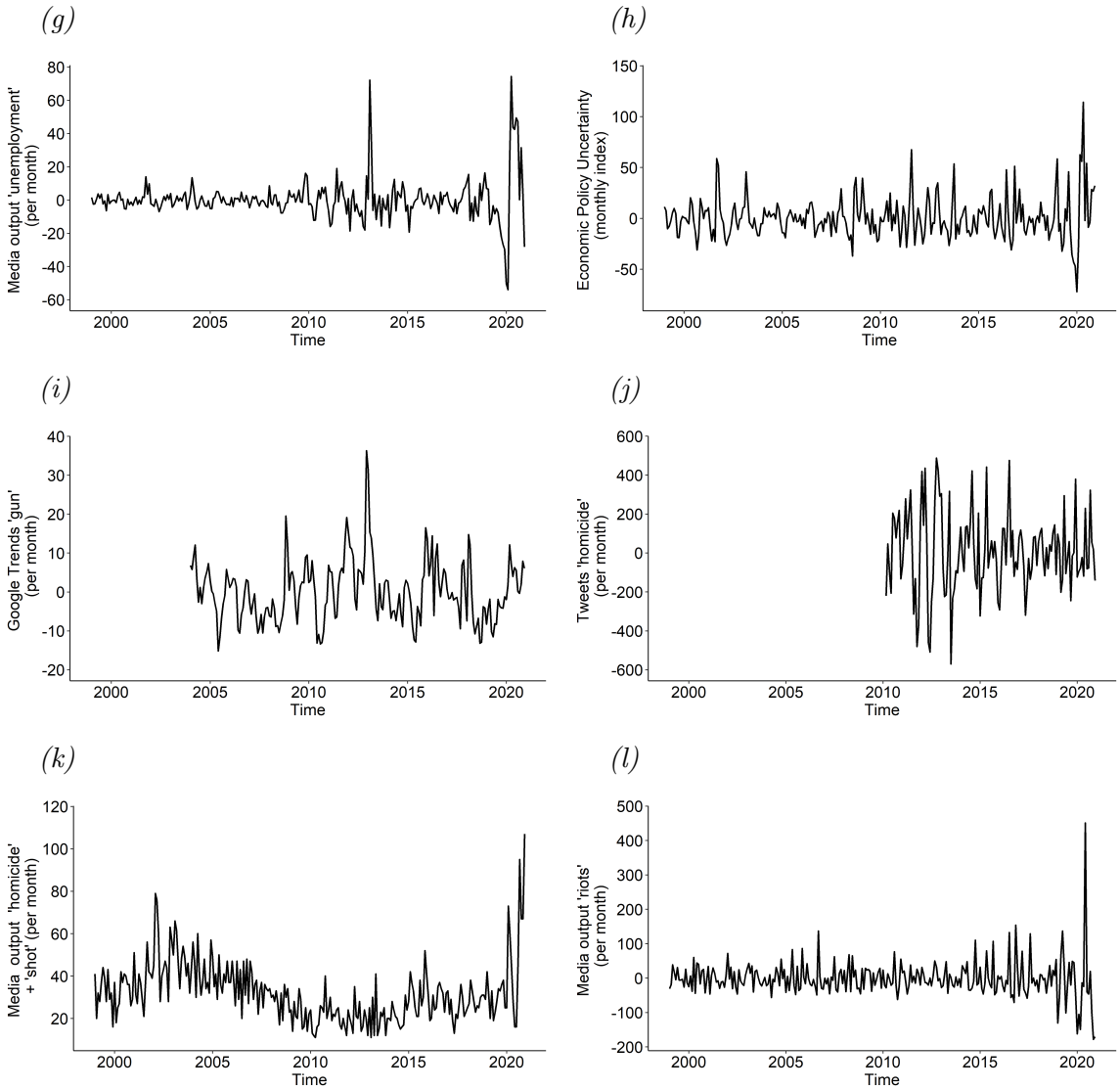
Notes. "Firearm" (red, •) refers to any type of gun (shotgun, rifle, handguns), "knife" (dark green, ▲) to cutting instruments, "personal weapons" (light green, ■) to the use of the body (hands, feet, etc.) as weapon, and "all other" (blue, +) to blunt objects, vehicles, and any other object used as a weapon.

Figure 4.2. Monthly time-series collected for the period from January 1999 to December 2020.



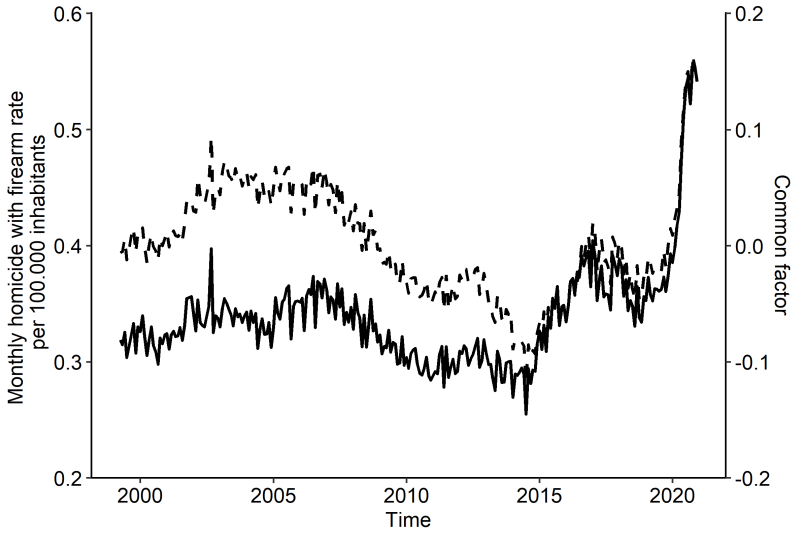
Notes. (a) Monthly homicides with firearm rate per 100,000 inhabitants from CDC (seasonally adjusted). (b) Provisional quarterly annualized estimate rate per 100,000 inhabitants for homicides with firearm from CDC. (c) Deaths in incidents involving guns from Guns Violence Archive (seasonally adjusted). (d) Monthly homicides rate for the aggregated cities of New York, Chicago, and Philadelphia (seasonally adjusted and detrended). (e) Monthly Google Trends for the word "Homicide" (seasonally adjusted and detrended). (f) Monthly background checks from NCIS (seasonally adjusted and detrended).

Figure 4.2. Monthly time-series collected for the period from January 1999 to December 2020 (continued).



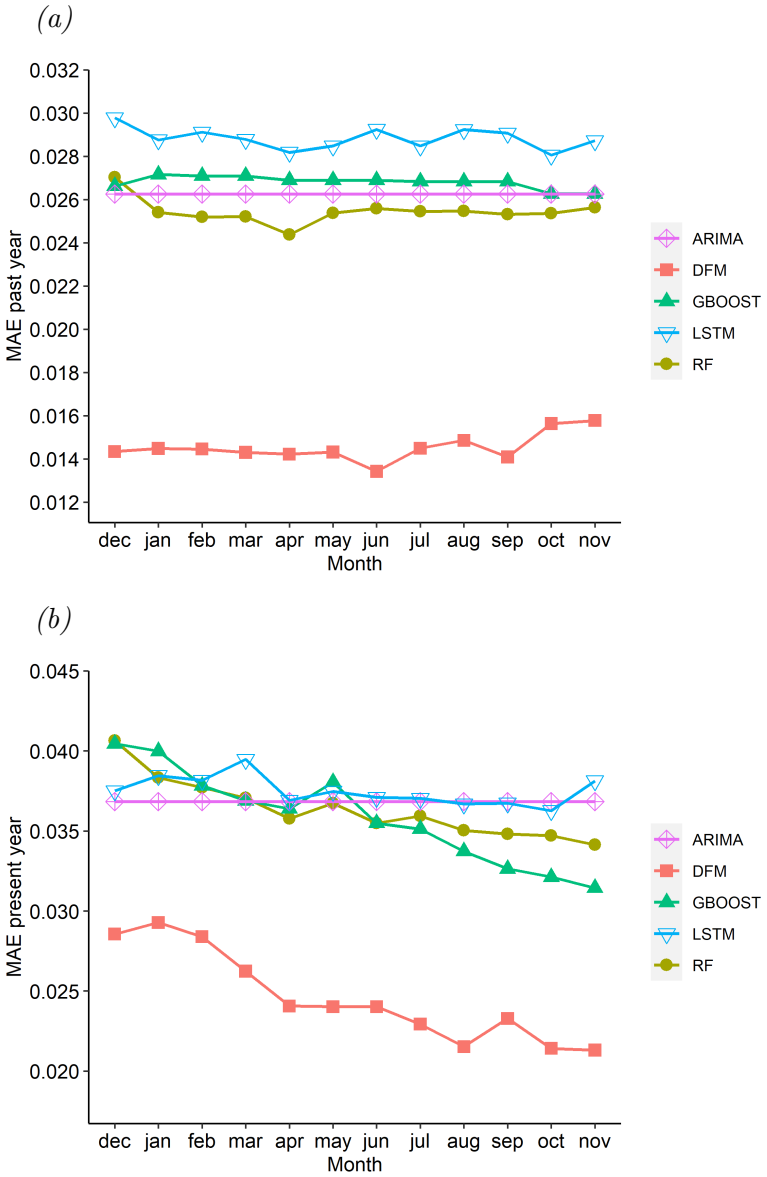
Notes.(g) Media output for the news containing the words "homicide" with "shot". (h) Media output for the news containing the word "riots" (seasonally adjusted and detrended). (i) Media output for the news containing the word "unemployment" (seasonally adjusted and detrended). (j) Monthly Economic Policy Uncertainty (EPU) index (seasonally adjusted and detrended). (k) Monthly Google trends for the word "gun" (seasonally adjusted and detrended). (l) Monthly geo-located tweets in the US containing the word "homicide" (seasonally adjusted and detrended).

Figure 4.3. Adjusted model fitted for the whole sample period from January 1999 to December 2020.



Notes. The dashed line represents the common factor (right vertical axis) and the black line the reconstructed series from the model on monthly homicides with firearm.

Figure 4.4. Out-of-sample pseudo real-time comparison between the DFM, AI models, and ARIMA for the period from January 2008 to December 2020, in terms of MAE: (a) previous and (b) present years.



Notes. Each point represents an MAE value for either the entire year preceding the month at which the prediction is made (a) or the entire year when the prediction is made (b). DFM (red, ■), RF (dark green, •), GBOOST (light green, ▲), LSTM (light blue, ▽), and ARIMA(3,0,3) (violet, ⊕).

Chapter 5

What drives the EMU business cycles connectedness: evidence from a TAR dynamic panel model

5.1 — Introduction

In recent years, the global economy has encountered economic shocks that have posed unprecedented economic challenges. Building upon the aftermath of the Great Recession, the Covid-19 pandemic, followed by the Ukrainian War, further exacerbated these challenges. Consequently, the economic effects of these shocks have spurred increased academic research in business cycle analysis. These shocks come at a time when the use of telecommunications has become widespread (Gomez-Barroso and Marbán-Flores, 2020), tourism has become a key activity worldwide (UNWTO, 2020) and trade has reached its highest levels (UNCTAD, 2022). Consequently, in this globalized economy, economic shocks are no longer an idiosyncratic problem to be studied in isolation; rather, they entail a certain degree of contagion and transmission to other countries (Sebestyen and Ilokskics, 2020). This enhances the need to investigate the role of globalization in the propagation of shocks, still unclear as highlighted by Kose, Prasad, and Terrones (2003), and Eickmeier (2007), to gain a deeper understanding of the interconnectedness and interdependence of economies.

Within the framework of this globalized and interdependent economy, the study of the transmission of shocks and the synchronization of business cycles assumes even greater significance within the European Union and the Economic Monetary Union (EMU), owing to their economic integration (Camacho and Perez-Quirós, 2006; Giannone, 2010). In this regard, Gehringer and König (2021) argue that, while monetary integration helped synchronicity, the debt crisis halted the tendency. Nevertheless, there is no consensus on the role of monetary integration in synchronisation, since previously Crespo-Cuaresma and Lopez-Amador (2013) for instance, contend that monetary integration did not enhance synchronization but instead fostered greater interdependencies among a group of countries. Therefore, the study of inter-

dependencies between EMU countries has assumed crucial importance in the current juncture, and a more in-depth analysis of them and their drivers has become even more relevant.

While there is no single methodology for studying the transmission of shocks and interdependencies, a convenient approach to analyze the transmission of shocks is the connectedness measure given by Diebold and Yilmaz (2009). Based on the Vector Autoregressive (VAR) model, the method focuses on the analysis of the forecast error variance decomposition (FEVD) to gauge spillovers and dependencies. This method, supposes two main advantages from other approaches to measure spillovers such as the inference through a Global VAR set up in Dees and Vansteenkiste (2007), the Minimum Spanning Tree in Matesanz et al. (2017), or the multilevel dynamic factor model approach in Camacho, Pacce and Perez-Quirós (2020). First, it provides a simple procedure to obtain interdependencies between economies based on a well-established econometric tool as the FEVD of a VAR model is. Second, its widespread adoption and popularity in recent years underscores its practicality, in contrast to recent alternatives such as the joint spillover index by Lastrapes and Wiesen (2021) (based on joint conditional forecasts), the model-free connectedness approach in Gabauer, Chatziantoniou, and Stenfors (2023) (which constrains Diebold and Yilmaz's proposal), or the PCMCI algorithm by Runge (2018), a time-lagged causal discovery framework based on conditional independence testing but whose relationships, while allowing for nonlinear relationships, are more difficult to interpret in terms of magnitude in a system.

Noteworthy applications of this measurement include Diebold and Yilmaz (2015) examination of the connectivity of business cycles from G-6, or the evaluation among the European Union, China, and the United States in Antonakakis, Chatziantoniou, and Filis (2016), and the analysis in Arčabić and Škrinjarić (2021) for EU countries. Albeit there exists a growing number of studies on connectivity in business cycles, there is a lack of analysis for the EMU, beyond that of Magkonis and Tsopanakis (2020), which, however, focuses on the banking sector and money markets. Nonetheless, its progressive construction process together with their common monetary policy but with a different exposition to the Ukrainian war and previous shocks, make their analysis a topic of interest.

Moreover, this definition of connectivity enables to evaluate different granularity levels, ranging from the measurement of connectivity between the business cycles of two specific countries, to a measure of global connectivity involving a set of countries. Through this method, not only a static analysis is possible but also a dynamic one from which we are able to identify periods of higher agreement. Moreover, the method also allows for an offset between the agents (in this case, countries), enabling asynchronous responses to each other's actions.

Nevertheless, to the best of our knowledge, no econometric models have yet been set up to ascertain the factors that determine interdependences. Related to this, Dees and Vansteenkiste (2007) rely on the variance decomposition of a Factor Structural VAR to explain linkeages into common, idiosyncratic and spillovers shocks, whereas Ngene (2021) explores the drivers of the

connectedness of the U.S equity sectors during different business cycles, and Crespo-Cuaresma (2022) analyzes the role of uncertainty in the synchronisation of business cycles. In this study we take a different aproach, based on the adaptation of the gravity model for trade, proposed first by Tinbergen (1962), and extended by Anderson and van Wincoop (2003), to the context of spillovers. This strategy entails a comprehensive exploration of interdependence between pairs of countries, encompassing considerations such as geographical distance and economic size. Additionally, we incorporated other bilateral data such as tourist and export exchanges, along with a measure of production similarity.

In addition, the persistence of the series and its susceptibility to fluctuations serve as indicators of potential temporal non-linear dynamics, which should be considered when modeling spillovers. Within the realm of nonlinear dynamic panel models, Kremer et al. (2013) develop the dynamic panel threshold model which utilizes the forward orthogonal deviations transformation. However, the approach requires the assumption of exogeneity for the variable that determines the non-linearity. In contrast, Seo and Shin (2016) propose an alternative approach based on first-difference GMM, where both the threshold variable and regressors are permitted to be endogenous. Building upon these findings, we propose a dynamic Threshold Autoregressive (TAR) panel model to elucidate the dynamics of the spillovers (obtained by applying the Diebold methodology to monthly industrial production as a proxy for the business cycle) employing the global connectivity index as the defining series characterizing the states of these dynamics. Subsequently, we conducted a rigorous comparative analysis of our proposed model against the temporal gravity model and its extension to a dynamic panel model estimated by GMM.

The empirical findings yielded several noteworthy insights. First, the estimation of spillovers revealed a high degree of integration between the countries in our sample, with Spain, Germany, and France as the main transmitters of business cycles (both in terms of their ability to originate and transmit shocks, as well as their ability to cope with external shocks). In addition, the connectedness is sensitive to the fluctuations in the business cycles, notably increasing during periods of economic downturns, aligning with findings observed Arčabić and Škrinjarić (2021). On top of that, the inferred network's topology (i.e. the structure of the network created by the interrelathionships) revealed different interaction groups, which also change while analyzing the topologies of other sectors where interaction between countries is important, such as tourism or exports. These communities, defy the conventional core-periphery view, aligning with outcomes reported in Matesanz et al. (2017). Second, outcomes derived from the extended gravity model with time dimension suggested that distance and differentiated GDP play an important role in the spillovers dynamics together with exports, tourism and industrial production similarity. Nevertheless, when extending the analysis to encompass a dynamic nonlinear panel approach, we observed a nuanced picture. Specifically, the previous time step level of the spillovers together with the exports and tourism are the ones that play a more important role in both upper and lower regimes, with the differentiated GDP being relevant

just in the upper regime and industrial similarity marginally just in the lower regime. This nonlinear extension also enables the creation of an indicator with two differentiated states with different levels of connectivity, with the upper regime indicative of high global connectedness.

These results hold significant implications and offer valuable guidance for policymakrs seeking to mitigate the effects of economics downturns period. In order to enhance the effectiveness of their measures, our results suggest a shift from country-specific strategies to measures common to countries with similar productive structure, together with a special care on tourism and exports. Moreover, the indicator derived from our model stands as a valuable tool for identifying strategic intervention points, where countries are more interconnected and thus where interventions are likely to yield the most impactful results

The rest of the chapter is organized as follows. Section 2 introduces the approach to compute spillovers and the adaptation of the gravity model to a nonlinear dynamic panel frame. Section 3 details a preliminary analysis of the data. Section 4 applies the approach to obtain the spillovers from EMU countries, analyzes the connectedness dynamics, and fits the panel data models to explain their dynamics. This section compares the estimation from the gravity model, the dynamic panel data alternative and our nonlinear dynamic proposal. Section 5 concludes and outlines some further research lines. This is followed by the references, appendices, tables and figures referred to in the chapter.

5.2 — Methods

5.2.1. Connectedness measurement

The measurement of connectivity proposed by Diebold and Yilmaz (2009) is based on the forecast error variance decomposition (FEVD) from a VAR model. In particular, FEVD specifies the proportion of the variability of a variable i of the VAR model which is explained by an unexpected shock in each of the N endogenous variables of the system. For our case, by stacking the monthly industrial production indexes from the Organisation for Economic Co-operation and Development (OECD) database in a vector Y_t, we can model it as a vector autoregression of order p, which in its structural form can be defined as

$$A_0 Y_t = \delta + A_1 Y_{t-1} + ... + A_P Y_{t-p} + \epsilon_t, \tag{C.1}$$

where, ϵ_t is a vector of i.i.d. shocks following $\mathcal{N}(0, \Sigma)$, with Σ their covariance matrix. One possibility is to estimate the FEVD from the structural shocks is by identifying the model in its reduced form after imposing restrictions, such as short run, long run or sign restrictions.

Another alternative is to compute the generalized forecast error variance decomposition

instead, proposed by Pesaran and Shin (1998), which is the one we adopted in our study. Without going into detail, this procedure does not assume a specific identification structure is assumed for structural shocks and Montecarlo techniques are used to estimate the unrestricted variance decomposition in the identification of shocks. Therefore is no need to estimate the parameters in the structural form of the model, and the elements of the covariance matrix of the forecast errors are obtained as

$$\Delta_{ij}^{h} = \frac{\sigma_{jj}^{-1} \sum_{r=0}^{h-1} (e_i' \psi_r \Omega e_j)^2}{\sum_{r=0}^{h-1} (e_i' (e_i' \psi_r \Omega \psi_r' e_i))}, \tag{C.2}$$

where ψ_r is the matrix that measures the response of the variables to a shock in the MA expression of the reduced form r periods ahead, e_j is a vector of zeros but in the $j-th$ position and σ_{jj} is the $j-th$ element in the diagonal of the Ω matrix of covariances. These elements are usually expressed normalized, so the columns of the matrix sum up to one each,

$$d_{ij}^{h} = \frac{\Delta_{ij}^{h}}{\sum_{j=1}^{N} \Delta_{ij}^{h}}. \tag{C.3}$$

The elements d_{ij}^{h} will form the matrix of the variance decomposition. Once the forecast error variance decomposition is obtained, connectivity at horizon h can be defined as the share of the year-on-year growth rate of a country's industrial production index that is explained by shocks in the other countries analyzed. Thus, the directional connectivity of country j to country i will be defined as

$$C_{i \leftarrow j}^{h} = d_{ij}^{h}, \tag{C.4}$$

so there will be $N^2 - N$ directional pairwise connectivities, and in general $C_{i \leftarrow j}^{h} \neq C_{j \leftarrow i}^{h}$. The net directional connectivity between two countries will be defined as

$$C_{ij}^{h} = C_{j \leftarrow i}^{h} - C_{i \leftarrow j}^{h}, \tag{C.5}$$

resulting positive when the business cycle of country i has a larger influence on the business cycle of country j than from j to i. On top of that, the net effect from the rest of the countries in a specific country can be obtained by adding the directional connectivities in each row but the elements in the main diagonal of the matrix

$$C_{i \leftarrow \cdot}^{h} = \sum_{j=1}^{N} d_{ij}^{h}, \quad j \neq i, \tag{C.6}$$

while one can also compute the connectivity "to others" adding up in each column but for the elements in the diagonal, this is how much country j affect the rest of the countries

$$C_{\cdot \leftarrow j}^{h} = \sum_{i=1}^{N} d_{ij}^{h}, \quad i \neq j, \tag{C.7}$$

from which we infer the ability of a country to transmit its business cycle to other countries. As a consequence, the net total connectivity of a country, this is the net transmission of the shocks of its business cycle to the rest of the countries in the sample will be

$$C_i^h = C_{\cdot \leftarrow i}^h - C_{i \leftarrow \cdot}^h, \tag{C.8}$$

with the sign of the expression determines whether a country is a net transmitter or a net receiver. The connectivity matrix is then adjusted to normalize the values of its elements, allowing them to be interpreted as percentages relative to the total prediction error variance, all while preserving the original connectivity ratios. The adjustment consists of dividing the connectivity by the number of countries in the set, in the same way as the global connectivity index has been calculated.

Finally, the total connectedness of the system, known as the connectedness index, will summarize to which extent countries in the sample influence into each other. The higher the index, the higher connectedness will result. It will be obtained by adding up the elements outside of the main diagonal divided by the number of countries in the sample (as in the transformation of the connectiviy matrix) and will vary from 0 up to 100, i.e. from a complete lack of spillovers to a complete explanation of shocks by spillovers. Additionally, connectedness can be also analyzed in a dynamic frame by estimating the model for a rolling window of a particular size, w, obtaining a different estimation of the model's parameters for each sample. So that, with T the sample size, we might get $T - w$ measurements of connectivity, allowing us to analyze its evolution in time.

5.2.2. Dynamic Panel Gravity modeling

Just as gravity models (Tinbergen, 1962) are employed in their application to trade, a similar rationale can be extended to spillovers. Thus, it would stand to reason that a country could be more influenced by another country if they are in close proximity (greater exchanges of human capital and goods capital), and this influence, in turn, could also be thought as a function of the size of the economy, since it would be reasonable for a larger economy to have a greater influence on a smaller economy. In this vein, the simpler version of the generalized gravity model of trade (Anderson and van Wincoop, 2003) can be adapted to the spillovers as

$$C_{j \leftarrow i} = \beta_0 \prod_k^K X_{j \leftarrow i,k}^{\beta_k} \epsilon_{j \leftarrow i}, \tag{C.9}$$

with $X_{j \leftarrow i,k}$ the K covariates that explain the bilateral connectedness, including bilateral data but also unilateral, and $\epsilon_{j \leftarrow i}$ i.i.d. errors. In the simpler setup, the gravity model will include GDP of each country of the repective pair together with the distance among them, measured as physical distance as in our case, or through alternatives measures (sharing borders, language

or colonial relation in the past). However, the inclusion of additional explanatory variables is widely accepted in the literature of gravity models (Bussière, Fidrmuc and Schnatz, 2008), and based on this we will add to the model bilateral data of different sectors.

On top of that, and due to data availability for the time span, the time dimension can be added to the model, which can be rewritten in logarithms as

$$lnC_{j\leftarrow i,t} = \beta_0 + \sum_{k=1}^{K} \beta_i lnX_{j\leftarrow i,k,t} + u_{j\leftarrow i}, \tag{C.10}$$

The traditional approach of gravity models is their estimation through GLS, however, Silva and Teneyro (2006) showed that this configuration can lead to problems in the very likely presence of heteroscedasticity, demonstrating that the estimation through Pseudo Poisson Maximum Likelihood (PPML) is more appropriate, and provides advantages over alternative methods such as feasible GLS or nonlinear least squares.

One element to be taken into account that has recently appeared in the literature is the fact that gravity models of trade tend to include dynamics into the panel-data model, once the high persistence of the series of trade has been accepted (Eichengreen and Irwin, 1997). This persistence results also to be the case for spillovers, so that by paralleling the reasoning, equation C.10 can be extended as

$$lnC_{j\leftarrow i,t} = \beta_0 + \beta_1 lnC_{j\leftarrow i,t-1} + \sum_{k=2}^{K} \beta_i lnX_{j\leftarrow i,k,t} + u_{j\leftarrow i}. \tag{C.11}$$

Nonetheless, the introduction of dynamics in the series creates an inconsistency problem in the estimation for fixed effects panel models (Nickel, 1981), and, as shown by Olivero & Yotov (2012) also for PPML estimations. To overcome this issue, De Benedictis and Vicarelli (2005), and Martinez-Zarzoso (2013), find that the system-GMM (Blundell and Bond, 1998) procedure is robust and preferred to the first-difference GMM from Arellano-Bond (1991), as the latter is not suitable in the presence of persistence.

Besides, the spillovers will be characterized by an asymmetric behavior, as business cycles do (Burns and Mitchell, 1946), thereby implying a source of uncertainty in accepting linearity in the models. In this vein, in the recent years the literature has analysed the problem of non-linearities in panel models. For instance, Hansen (1999) proposes a static panel threshold model where the coefficient can vary depending on the value of an exogenous variable. Based on this analysis, and to allow for endogenous regressors, Kremer et al. (2013) combined the forward orthogonal deviations transformation by Arellano and Bover (1995) and the instrumental variable estimation. However, to avoid the assumption of exogeneity on the threshold variable, Shin and Seo (2016) developed a threshold autoregressive model estimated through first-differenced GMM, in which both regressors and the threshold variable are allowed to be

endogenous, so that

$$lnC_{j\leftarrow i,t} = (1, X^*_{j\leftarrow i,k,t})\beta_1 1(q_{it} \leq \gamma) + (1, X^*_{j\leftarrow i,k,t})\beta_1' 1(q_{i,t} \geq \gamma) + u_{j\leftarrow i}. \tag{C.12}$$

where $X^*_{j\leftarrow i,k,t}$ the covariates including also the lagged connectivity in our case, and $q_{i,t}$ is the threshold variable at time t which in our case is the inferred global connectivity index. Based on this approach, a TAR model (Tong, 1990) can be considered to be applied where both the intercept and coeffients changes according to the threshold variable. Thus, for this proposal any possible problem of endogeneity of this treshold variable and the spillovers will not result into a problem as in the other approaches would be. Through grid search the threshold γ that minimizes the mean square error of the fitting through pooled least squares estimation can be obtained. This approach yields an indicator function of global connectivity, $I(q_{it} < \gamma)$ with q_{it} the global connectivity index, for the entire time dimension of the panel, indicating one state with a high connectivity and another state with low connectivity for the panel.

5.3 — Data

To characterize the business cycles, we chose the monthly industrial production indexes from the OECD database. The frequency of publication of Industrial production is higher compared to Gross Domestic Product, enabling a larger number of observations and a better gauge of the economic cycle. This is especially significant in the case of short-lived recessions, such as the one experienced during the health crisis.

Specifically, we use their year-on-year rates for the period from January 1966 to November 2022 for eleven of the eurozone countries with a longer period of data available: Austria, Belgium, Finland, France, Germany, Greece, Italy, Luxembourg, The Netherlands, Portugal, and Spain. During this period, there will be fluctuations both with positive growth and contractions. By utilizing a year-on-year rate, we can overcome issues related to stationarity and seasonality that are commonly encountered in industrial production series.

To get a picture closer to economic reality, a systematic outlier detection process was performed. In particular, we considered as outliers any of the observations identified as such by Isolation Forests (Liu et al., 2008) with a larger anomaly score than 0.9, Local Outlier algorithm (Breunig et al., 2000) with a larger factor than 3, or DBSCAN (Ester & Sander, 1996) with a distance larger than 3.5 (around 10% of the average range), with this parameter making reference to the maximum distance that two points can be from one another while still included in the same cluster, and with at least two points in the cluster of outliers. Through this preprocessing, we take into account both global outliers and local outliers. For the COVID-19 period, we just adjusted the year-to-year rates from March to August 2021, to avoid a

sudden increase after a sudden decrease that might affect autoregressive models estimation. The outliers points obtained in the process were filled in using a Kalman filter procedure for univariate ARIMA models (Hyndman and Khandakar, 2008; Moritz and Bartz-Beielstein, 2017). Although the Dicky Fuller tests rejected the presence of unit roots, which is required for the estimation of the VAR model, the series presented different outliers peaks (due to, for example, wars or turmoil).

To elaborate the explanatory model of the spillovers, we first gathered data from GDP, which was monthly linearized, and we defined the distance between pairs of countries as the physical distance between the capital and the closest border. Nevertheless, despite the availability of these two variables for the same time span as industrial production, the inclusion of three more additional explanatory variables can only be implemented starting from the year 1995 since bilateral Eurostat data became available from that year onward, restricting the analysis to that period. First, tourist exchange data were obtained from Eurostat's monthly series of inter-country air passengers. This would be, in some cases of missing values, extended back to 1995 using the dynamics of the series of nights by country of origin, available from Eurostat from 1995 to 2011. Second, bilateral export data were collected using the EU trade series by Eurostat's BEC by partner, available since since 1988. Third, we created a measure of the similarity of the industrial production. To this end, we reconstruct the composition of countries' production as follows. Firstly, we obtained the production weights for the countries as a percentage of production of intermediate goods, durable goods, non-durable goods, capital goods, and energy, published for the year 2015. Then, using the dynamics of each of these productions from Eurostat, we reconstructed the weights for the period up to 1995 and forward to 2020. At this stage, we computed the cosine similarity between each pair of countries at time t based on the industrial composition for each country at time t, $ic_{i,t}$, calculated as the inner (or scalar) product of the two vectors,

$$cosine_{ij,t} = \frac{\sum_{n=1}^{5} ic_{i_n,t} ic_{j_n,t}}{\sqrt{\sum_{n=1}^{5} ic_{i_n,t}^2 \sum_{n=1}^{5} ic_{j_n,t}^2}},$$

where n represents the five groups of production. Although already described in Appendix A of Chapter 3, it should be recalled that the similarity can range from -1 to 1, where -1 represents total dissimilarity, zero represents no similarity and 1 complete similarity. As can be noted from Figure 5.1, countries such as France and Germany present a more similar industrial composition than for instance Greece and Germany, as Greece is a country whose economy is mainly based on tourism and its production therefore is focused on.

5.4 — Results

5.4.1. Connectedness and network structure

For the VAR model estimation from which we will obtain connectivity, we followed the Schwartz criteria, with the configuration of $p = 2$ yielding the optimum setting (SC: 25.66) with the whole sample. The connectivity table is usually transformed to relativize the value of each of the elements so that they can be interpreted as the percentage they represent over the total variance of the prediction error. The transformation involves dividing the connectivities by the total number of variables, utilizing the previously computed connectedness index. Tables 1 and 2 present the static connectedness during the short-term and long-term, specifically two and twelve months following the unexpected shock.

In the short run, the total connectedness index is 36.69%, unveiling that more than one-third of the forecast error variance is due to shocks transmitted between countries, while 63.31% is due to shocks from the countries themselves. As can be expected, the elements in the diagonal are the largest in the matrix, since they represent to which extent a shock with origin in one country affects that country. Nevertheless, there are certain significant relations. For instance for a 2-month horizon, while Italy has a 1.26 connection to France, the latter has a connection of 1.15 to Spain, and Germany a connection of 1.13 to Austria. The opposite direction relations are smaller, with France having a connection to Italy of 1.06, Spain to France of 1.03, and Austria to Germany of just 0.50. As a summary in a visual shape, the resulting whole network representation with significant connections (larger than 0.05) is depicted in Figure 5.2. To ensure that the number of parameters to be estimated in the VAR model does not pose an estimation problem in our case, we conducted a robustness test. We examined the accuracy of the inferred configuration using the GNAR methodology introduced by Knight et al. (2020), which analyzes network dynamics. The proposed configuration was among the tested options that exhibited lower fitting errors (see Appendix A for more details).

On its part, the total connectedness index for the long term increases up to 56.85%, meaning that more than half of the forecast error variance is due to shocks transmitted between countries. On top of that, this result shows that perturbations are still being transmitted after several months. In this case, the connection from Italy to France rises up to 1.31 and 1.12 in the opposite direction, while from Spain to France is 1.49 (1.13 in the opposite direction), and 1.51 from Germany to Austria (decreases to 0.36 in the opposite direction).

Moreover, the net receivers and transmitters from the countries in the sample can be obtained. From Tables 3 and 4, Austria, Belgium, Finland, Greece, The Netherlands, and Portugal are the receiver countries, while Germany, Italy, and Spain are the transmitter countries. For the case of Luxembourg and France, in the short run is a receiver but in the medium term

turns to be a transmitter. In addition, we can suspect that COVID-19 pandemic increased the role of Italy as a transmitter, as it was the country where the virus appeared first among the countries in the sample. This is actually confirmed by performing the analysis up to December 2019, reflected in Table 5 which shows the differences at horizon $h = 2$ between both cases. Indeed, the large positive difference of 3.26 in the case of transmission for Italy, larger than the difference in reception (1.10), although still smaller than its net connectedness with the whole sample (2.41 in Table 3), represents a significant amount of the transmission estimated, showing that, effectively, the pandemic intensifies Italy's classification as a transmitting country. For the rest of the countries, their nature also did not change.

Once a certain topology has been inferred, descriptive measures of the system can be obtained, as well as analyzing areas of closer interaction. In terms of the former, and identifying an edge as a connection from a country to another and a node as a production indicator time series of a country, the main measures that describe the structure of a network are: the density of edges per node (as the quantity of spillovers that a country takes part of), the clustering coefficient (as the proportion of connections among the neighbours of a node which are actually take place compared with the number of all possible connections), the average path length (as the mean quantity of the inverse of the connections), the diameter of the network (understanding distance as the inverse of the weight of an edge) as the maximum distance, the mean degree (to analyze the average number of edges that arrive to an average node) and reciprocity (to study the proportion of connections bidirectional between two nodes). In our case, to obtain the main features of the topology inferred, only edges that accounted for more than 5 percent of a node's connectivity were retained to discard irrelevant connections and analyze network characteristics. The network with two months horizon had a density of 0.43, a diameter of 0.45, a reciprocity of 0.58, and an average path length of 0.19, indicating a network with a high density of edges, with nodes close together, and largely bi-directional edges. Besides, the features indicate indeed that the nodes (or countries) are largely influenced by the other nodes: if the nodes were mostly explained by their own behavior, the number of edges with a connection of more than 5 % would be small.

Also very informative about the nature of a network structure, is the detection of communities in terms of interactions, where a community is understood as a set of nodes (i.e. industrial productions in our case) that are more densely connected to each other than to nodes outside of the community. Regarding the partition of the network, one of the most popular algorithms to detect communities in a graph is the Louvain algorithm (Pujol, Erramilli, and Rodriguez, 2009). In particular, the algorithm optimizes the quality of the partition of the network by maximizing iteratively the modularity, a measure of the strength of a particular partition of the network which is defined as

$$Q = \frac{1}{2m} \sum_{ij} [C_{j \leftarrow i} - \frac{(C_{\cdot \leftarrow i} + C_{i \leftarrow \cdot})(C_{\cdot \leftarrow i} + C_{j \leftarrow \cdot})}{2m}] \delta(c_i, c_j), \tag{C.13}$$

where $C_{j\leftarrow i}$ is the edge weight (connection) between nodes (countries in our case) i and j, m is the sum of all the weights in the network, c_i and c_j are the communities assigned to the nodes, δ is Kronecker's delta function, and where the numerator of the fraction represent the product of the degree of both nodes. The algorithm of optimization is based on initially assigning a community to each node, for which it is analyzed how much modurality will be increased when moving the node to another neighbor community (in the initial case that assigns a community to every node, these correspond simply to the neighbors of each node), choosing the configuration that increases it the most. Thus, a new graph is created with the nodes as the communities previously found, defining the weights (i.e. the connections) between communities as the sum of the weights of the nodes in the respective communities. Through iteration between both steps, the optimum membership in communities that maximizes modularity can be obtained.

In terms of communities, Figure 5.3 shows how, according to the Louvain algorithm applied through the *multinet* package (Magnani, Rossi, and Vega, 2021), exterior countries are more isolated from the rest of the EMU countries, such as Finland and Greece, and which are assigned as isolated communities, while two large groups are obtained, one characterized by Austria, Spain, Italy, and Germany, and the other by the Benelux countries, Portugal and France, the latter being the closest country to serve as a link between communities. These clusters found do not map the core-peripherical view (Lehwald , 2013), in line with the results also obtained in Matesanz et al., (2017). [1].

In addition to knowing that the connectedness is larger in the long term as obtained, it is a matter of interest to analyze up to which period this is the case, since at a certain period the forecast error variance has to stabilize. Figure 5.4 depicts the global connectivity index up to a horizon of 18 months. Two insights stand out. First, the shocks in the errors do not decrease with time in this system. Second, after around 12 months, the connectedness remains around 60%, increasing in that period by around 15% although most of the connectivity is therefore in the short term .

Once the connectivities have been obtained and their nature has been analysed, the dynamic connectedness was explored for a two months horizon through a VAR(2) recursively estimated. Figure 5.5 shows the connectedness index, with a rolling window of 60 months and a forecast horizon of two months ahead, together with the chronology established by the Euro Area Business Cycle Network Committee (EABCN). As can be noted, the connectedness seems to follow a cyclical path, with the caveat that the dynamic window consistently encapsulates data from a specified timeframe, and in order to effectively represent a transition, a requisite quantity of observations must manifest such a transition (which consequently introduces a

[1] A comparison of the network features and of the communities obtained in the production network with the networks obtained for two other variables with presumed cross-country influence as tourism and exports, shows that the production network is more similar to that of tourism than to that of exports, although in all cases the communities obtained are different, separating Mediterranean countries in the first case and traditionally exporting countries in the second (for more details, see Appendix B).

temporal lag before the index accurately portrays it). During economic recessions, the index uses to first shortly decrease, followed by a large increase during and right after the period, while in expansions the index turns to decrease after the mentioned increase that follows the recession to later keep stay around a lower level of connectivity, pointing to a certain nonlinear behaviour which will be discussed in the following subsection. In addition, it can be noted that the influence of globalization is notable from 1990, creating an increasing trend on the connectivity index, which remained stable up to the Great Recession, following again from that point the mentioned pattern, sensitive to the cycle. The sanitary crisis consequence of COVID-19 resulted in an unprecedented connectedness, with a peak of up to 80.

For the sake of the robustness of this insight, we performed the analysis with windows of different sizes (always with a two months ahead horizon) and obtained the mean difference between the peaks and the droughts and vice versa. The results summarized in Table 6 confirmed the dynamics and the sensitivity to the business cycles of the connectedness index. The level of the index remained around 50 at every window and the mean difference from the end to the beginning of recessions always stay high, with a sligh decline always in the recessions.

5.4.2. Explanatory models

Once the spillovers were estimated and their structure characterised, at this juncture we fitted the gravity model through PPML estimation with the addition of other covariates to obtain an explanatory view of the spillovers. The model was implemented with the inclusion of the time dimension, encompassing also a comprehensive range of variables.[2] Specifically, bilateral tourism, exports, and the computed similarity of industrial production were incorporated into the model together with GDP and distance. While spurious correlations typically poses smaller concerns in gravity models than in time series, Kao and Chiang (2000) showed that fixed effects estimators in panels that do not take stationarity into account might be biased. As a consequence, we conducted Panel Unit Root tests to check whether any covariate must be differentiated. The results of the four tests (Levinlin, IPS, Maddala-Wu, and Hadri), are presented in Table 7, confirming that the GDP series in logarithms (as present in our modeling) should be differenced, while the remaining variables did not require such transformation.With the established set up, the results presented in the first column of Table 8 show, for the unbalanced dataset from 1995, that all the variables were significant to explain the spillovers but bilateral exports and the differentiated GDP series. While the sign of the coefficent for distance was negative, it was observed a positive relation in the similarity measure, tourism, differentiated logarithm of GDP, and exports.

[2]In order to gain certainty of the need of including other variables out of those included in the simpler gravity model of trade such as GDP and distance, we performed an estimation of this simpler model through the inclusion of possible unobserved structures in the model, following Chen, Fernández-Val and Weidner (2021), from which it was cofirmed that more variables needed to be taken into account since some hidden structures were present. For further details, see Appendix C.

Nonetheless, due to the persistent nature of the connectivities (a feature that can be noted for instance in Figure 5.6 for the cases of the spillovers from Germany to France and from Greece to Germany), we employed a dynamic panel model in our analysis. In this vein, Roodman (2009) highlighted the possibility of weak instruments when using the Arellano-Bond estimator in dynamic panels, and potential limitations in small sample properties with the Blundell-Bond estimator. Thus, to address potential issues of weak instruments, we utilized a system-GMM procedure with lagged spillovers from two to five periods as instruments since our dataset was larger than the number of individuals in the panel. Upon incorporating the series into the dynamic panel model through the *plm* package, where we specified a fixed effects model and a two step GMM estimation, we observed that the autoregressive spillover term was a significant variable. On its part, distance and tourism remained statistically significant, as reflected in the second column of Table 8. Additionaly, the incorporation of autoregressive dynamics in the model rendered the differentiated GDP in logarithms for the sender country as a significant variable (exports and the differentiated GDP of the receiver country in logarithms also remained non-significant in this case, changing to also non significant coefficients both exports and the similarity in industrial production). To evaluate the presence of autocorrelation and the correct introduction of the autoregressive term, we performed the Arellano autocorrelation test. The test correctly rejected autocorrelation at lag one (z=-5.25, p-value<0.01) and failed to reject it at two periods (z=1.43, p-value=0.15) as suggested by the authors to declare absence of autocorrelation. Furthermore, the Sargan test indicated that the instruments were uncorrelated with the error term and supported their proper exclusion from the equation, as the test did not reject the null hypothesis of valid instruments ($\chi^2(1659)$=83.27, p-value=1.00).

Morever, a key aspect in the analysis is the presence of asymmetric behavior in connectivity and spillovers, which led us to explore the possibility of extending the model and fitting a non-linear dynamic model. To this end, we employed a TAR model as the natural choice, where the threshold variable is the global connectivity index previously estimated with a rolling window of 60 months. The estimation is done through first difference GMM, and in this case, both the intercept and the slope coefficients change according to the global connectivity index as the threshold variable, and was built based on the *dptee* package. The procedure used to infer the state at each of the periods is based on minimizing the error through a grid search. The instruments chosen for the estimations where the lagged series of the dependent variable up to five lags.

The estimated threshold was determined to be 54.20 and significant. In the high state, the autoregressive spillover term was found to be significant, along with the differentiated GDP in logarithms of the sender country, tourism (all with negative sign), and exports (with positive sign), results reflected as Model 3 in Table 8. In the low state, the autoregressive spillover term was found to be significant (with positive sign), together with exports, tourism and, marginally, industrial production (all with negative sign). The intercept, which also accounted for the fixed effect of distance, was estimated to be -26.71 (significant) in the upper state (and set to zero

in the lower state). As can be noted from Figure 5.7 the low state corresponds to one with a lower global connectedness index, while the high state to one with a high index. In particular, the financial crisis and the Covid period are included in the latter together with some short perios in 2018, 2003, and 2004. Moreover, all the variables but tourism changed their sign from state to state. With respect to that, the latter is reasonable. Tourism usually decreases in recessionary periods, where high state is mainly present and where connectedness tend to increase, and usually increases in expansion periods, where connectedness then to stay low or even decrease. With respect to the positive sign of exports in recessionary periods, it is a known fact that net exports are countercyclical (Backus, Kehoe, and Kydland, 1992), so that gain weight in economies in recessionary periods, when internal demand looses strength but they are not adversely affected as pointed in Abiad, Mishra, and Topalova (2014), although we would expect a positive sign in the low regime, although this is lower and not significant at higher confidence levels. .

Finally, to check the validity of the specifications obtained and the correct introduction on non-linearity, tests for the null of no threshold effects and for the validity of the overidentifying moment conditions were performed. For the former, Seo and Shin (2016) provide a test based on the supremum type statistics, which although they do not follow standard asymptotic distributions, its critical values can be found by boostrap procedures. In particular, in our case the Wald type supremum linearity test was performed, based on the bootstrap algorithm in Seo, Kim, and Kim (2019), whose result presented a small bootstrap p-value $< 10^{-2}$, which suggests strong evidence in favor of threshold effects and supporting the nonlinear model adopted. For the validity of instruments, the overidentifying restrictions test (or J-test) of exogeneity of the instruments, yielded a value of 192.77 and resulting in a high p-value, indicating that the null hypothesis of valid instruments could not be rejected, so that it can not be ruled out that the lagged series of the dependent variable up to five lags are exogenous instruments. Therefore, the TAR approach is found as the valid approach in our comparison, unveiling a differentiated behaviour of the spillovers according to the different degree of connectedness in expansionary and recessionary periods. This result is in line with the implementation in times of recession of economic transfer mechanisms such as the Support to mitigate Unemployment Risks in an Emergency (SURE) mechanism in 2020, the establishment of spending rules in 2011 or economic bailouts such as those of Greece in 2010 and Portugal in 2011, initiatives that unify the economic response in such periods, even if shocks hit countries asymmetrically.

5.5 — Conclusions

The magnitude of the last shocks in the world economies has increased the interest in the literature about the dynamics of the business cycles. Within a global economy, the mechanism of transimission of shocks throughout economies is a feature that needs to be taken into account

in this analysis. Besides, this is a key feature in the EMU, where the monetary union and the free movement of capital, goods, services, and people, create a framework of highly connected economies.

In this study, we analyze to which extent the business cycles of eleven countries of the EMU are connected and we propose and explanatory model of these spillovers. Initially, through the definition of connectedness of Diebold and Yilmaz (2009) based on the forecast error variance decomposition of a VAR model, we quantify the level of connection between the countries during more than 50 years. Moreover, the relations obtained were confirmed through a GNAR estimation for robustness. The study yielded five results to mention in terms of the spillovers found. First, the EMU countries analyzed present a high degree of connectedness. This interdependency is not just in the short run, but even increases in the medium term with the system stabilizes its global connectivity around 60 % after eight too twelve months, although the larger share takes place in the short run.

Second, from the analysis of connectedness' directionality, the key role of geography is revealed. While Spain, Italy, Germany, Luxembourg, and France are net transmitters of business cycles, Finland, Portugal, Greece, Austria, Belgium and, The Netherlands are net receivers. Third, and related to the above, the magnitude of Covid-19 pandemics and the arrival of the virus in first place to Italy among the countries in the sample, made this country to change from net receiver to net transmitter of perturbations. Fourh, the cluster analysis of the network unveiled two main community groups, which rebut the core-peripheric view, in line with the recent findings in Matesanz et al. (2016) in their synchronicity analysis..

Fifth, regarding the dynamics of connectedness, the index of global connectivity is not symmetric. We unveiled increases during recessions, resulting in a more compact response to more complicated situations, and less connection during expansions. This result, which is in line with Arčabić and Škrinjarić (2021), was also checked with analysis with different rolling windows. On top of that, the globalization might be the cause of a more interconnected system. However, the creation of the European Central Bank and the establishment of the euro as the single currency do not seem to increase dramatically the interdependencies of the business cycles for the countries in the sample. This result is in line with the finding of Giannone et al. (2010) who claimed that the adoption of the euro did not derive into any divergence or convergence on business cycles economies. In fact, we detected two differentiated communities and two isolated countries as optimizing partitions of the network.

Finally, once the topology of the network was inferred, a gravity model was fitted in a dynamic version to analyse and explain the spillovers, resulting significant variables such as industrial similarity, bilateral tourist flow, distance and GDP on a sample from 2000 to 2022. Nevertheless, to model the persistent dynamics of the spillovers, the gravity model was augmented to a dynamic panel model, for which the Arellano test discarded the presence of autocorrelation in the dynamic set up. On top of that, a non-linear set up was adopted, which was

supported by the supremum Wald test supporting the presence of threshold effects. Therefore, a TAR model whose transition variable was the global connectivity index was fitted. From this latter analysis an indicator was developped, characterised by two regimes with higher and lower interconnectedness. In particular, recession periods are included in the upper regime together with other periods with an observed smaller peak of conenctedness. Tourism, exports, and the autoregressive dynamics were found significant in both regimes, although in the case of exports the magnitude of the relation was found less intense in the lower regime, and the differentiated logarithm of GDP of the sender country was significant just in the upper regime. These results have important implications in terms of policy recommendations. First, it shows that countries' interconnectivity behaves differently in different periods, where the role of different bilateral variables changes. Second, policies will be easier to implement in times of high connectivity such as recessions and where the role of exports is important.

The study is not free of limitations. One of the major limitations of the analysis, namely the availability of variables for the dynamic panel exercise with a longer time sample. Thus, being able to extend the time series would allow a larger number of periods with different economic circumstances to be analysed and conclusions to be drawn with greater certainty. At this point, several different courses of action can be adopted as extensions. The results of the flows of perturbations in the business cycles allow to analyze statistically possible more complex networks with interactions not being taken into account here. In addition, the study of other variables might help to get a more detailed picture of the spillovers. In this vein, an analysis of the balance of payments could be complementary to the analysis. In particular, Diebold and Yilmaz (2015) reveal a generalized relationship between a country's trade surplus and its consideration as a net recipient of cycles (and vice versa). Closely related to this, recently Barigozzi (2022), extended the measurement of spillovers to a multilayer network, so that the analysis could take into account different networks for different sectors able to also allow inter-network connections. Other alternative might be trying to adapt the dynamic panel methodology to the modelling of a changing regime determined by a Markov chain could be a natural extension of the non-linear modelling proposed in this chapter. Last but not least, it must be noted that a country's transmission capacity does not control whether a country distributes a shock or whether a country is the first to receive a common shock exogenous to all countries that ends up reaching all countries in the group. Extending the connectivity measure by controlling for exogenous shocks as in a Global VAR could be an alternative to pursue.

5.6 — References

Abiad, A., Mishra, P., and Topalova, P., 2014. How does trade evolve in the aftermath of financial crises?. *IMF Economic Review*, 62, 213-247.

Anderson, J. E.,and van Wincoop, E., 2003. Gravity with Gravitas: A Solution to the Border

Puzzle *American Economic Review*, 93, 170-192.

Antonakakis, N., Chatziantoniou I.,and Filis G., 2016. Business Cycle Spillovers in the European Union: What is the Message Transmitted to the Core? *The Manchester School*, 84(4), 437-481.

Arčabić, V., and Škrinjarić, T., 2021. Sharing is caring: Spillovers and synchronization of business cycles in the European Union. *Economic Modelling*, 96, 25-39.

Arellano, M., and Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2), 277-297.

Arellano, M., and Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29-51.

Backus, D., Kehoe, P. J., and Kydland, F., 1992. Dynamics of the Trade Balance and the Terms of Trade: The S-curve. Federal Reserve Bank of St. Louis Working Paper 9211.

Barigozzi, M., Cavaliere, G., and Moramarco, G., 2022. Factor network autoregressions. arXiv preprint arXiv:2208.02925.

Blundell, R., and Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1), 115-143.

Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J., 2000. LOF: identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 93-104.

Burns, A. F., and Mitchell, W. C., 1946. Measuring business cycles (No. burn46-1). National Bureau of Economic Research. Cambridge, MA

Camacho, M., Perez-Quiros, G. and Saiz, L., 2006. Are European business cycles close enough to be just one?, *Journal of Economic Dynamics and Control* 30, 1687–1706.

Camacho, M., Pacce, M., and Pérez-Quirós, G. (2020). Spillover effects in international business cycles. Banco de España, Working Paper No. 2034.

Chen, M., Fernández-Val, I., and Weidner, M., 2021. Nonlinear factor models for network and panel data. *Journal of Econometrics*, 220(2), 296-324.

Crespo-Cuaresma, J., and Fernández-Amador, O., 2013. Business cycle convergence in EMU: A first look at the second moment. *Journal of Macroeconomics*, 37, 265-284.

Crespo-Cuaresma, J., 2022. Uncertainty and business cycle synchronization in Europe. *Applied Economics Letters*, 29(11), 1047-1053.

De Benedictis, L., and Vicarelli, C., 2005. Trade potentials in gravity panel data models. *The*

BE Journal of Economic Analysis & Policy, 5(1),

Dées, S., and Vansteenkiste, I., 2007. The transmission of US cyclical developments to the rest of the world. ECB Working Paper Series No 798.

Diebold, F.,and Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. *Economic Journal*, 119, 158-171.

Diebold F.,and Yilmaz K., 2015a. Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring. Oxford University Press.

Eichengreen, B., and Irwin, D., 1997. *The role of history in bilateral trade flows.* National Bureau of Economic Research working paper 5565. Cambridge, MA.

Eickmeier, S. 2007. Business Cycle Transmission from the US to Germany-A Structural Factor Approach. *European Economic Review*, 51: 521–551.

Ester, M., Kriegel, H. P., Sander, J., and Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the2nd International Conference on Knowledge Discovery and Data Mining. Portland: AAAIPress, Vol. 96, No. 34, pp. 226-231.

Gabauer, D., Chatziantoniou, I., & Stenfors, A., 2023. Model-free connectedness measures. *Finance Research Letters*, 54, 103804.

Gehringer, A., and König, J., 2021. Recent patterns of economic alignment in the european (monetary) union. *Journal of Risk and Financial Management*, 14(8), 362.

Giannone, D., Lenza, M. and Reichlin, L., 2010. Business Cycles in the Euro Area. *Europe and the Euro.* Ed. Alesina, A. and Giavazzi, F., 141–167. The University of Chicago Press.

Gómez-Barroso, J. L., and Marbán-Flores, R., 2020. Telecommunications and economic development–The 21st century: Making the evidence stronger. *Telecommunications Policy*, 44(2), 101905.

Hansen, B. E., 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics*, 93(2), 345-368.

Hyndman, R. J., and Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 27, 1-22.

Jeffreys, H., 1948. *Theory of probability.* ClarendonPress. Oxford, 2nd edition.

Knight,M.,Leeming,K.,Nason,G.,and Nunes, M., 2020. Generalised network autoregressive processes and the GNAR package. *Journal of Statistical Software*, 96, 1-36.

Kose, M. A., Prasad, E. S., and Terrones, M. E. (2003). How does globalization affect the synchronization of business cycles?. *American Economic Review*, 93(2), 57-62.

Kremer, S., Bick, A., and Nautz, D., 2013. Inflation and growth: new evidence from a dynamic panel threshold analysis. *Empirical Economics*, 44, 861-878.

Lastrapes, W. D., and Wiesen, T. F., 2021. The joint spillover index. *Economic Modelling*, 94, 681-691.

Lehwald, S., 2013. Has the Euro Changed Business Cycle Synchronization? Evidence from the Core and the Periphery. *Empirica* 40(4): 655-684.

Liu, F. T., Ting, K. M., and Zhou, Z. H., 2008. Isolation forest. 8th International Conference on Data Mining, 413-422. IEEE.

Magnani, M., Rossi, L., and Vega, D., 2021. Analysis of multiplex social networks with R. *Journal of Statistical Software*, 98, 1-30.

Magkonis, G. and Tsopanakis, A., 2020. The financial connectedness between eurozone core and periphery: a disaggregated view. *Macroeconomic Dynamics*, 24(7), 1674-1699.

Martínez-Zarzoso, I., 2013. The log of gravity revisited. *Applied Economics*, 45(3), 311-327.

Matesanz Gomez, D., Ferrari, H. J., Torgler, B., and Ortega, G. J. (2017). Synchronization and diversity in business cycles: a network analysis of the European Union. *Applied Economics*, 49(10), 972-986.

Moritz, S., and Bartz-Beielstein, T., 2017. imputeTS: time series missing value imputation in R. The R Journal, 9(1), 2017.

Ngene, G. M., 2021. What drives dynamic connectedness of the US equity sectors during different business cycles?. *The North American Journal of Economics and Finance*, 58, 101493.

Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica: Journal of the econometric society*, 1417-1426.

Olivero, M. P., and Yotov, Y. V., 2012. Dynamic gravity: endogenous country size and asset accumulation. *Canadian Journal of Economics/Revue canadienne d'économique*, 45(1), 64-92.

Pesaran, H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate model. *Economics Letters*, 58, 17-29.

Pujol, J. M., Erramilli, V., and Rodriguez, P., 2009. Divide and conquer: Partitioning online social networks. *arXiv preprint* arXiv:0905.4918.

Roodman, D., 2009. How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*, 9(1), 86-136.

Runge, J., 2018. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7).

Sebestyen, T., and Iloskics, Z., 2020. Do economic shocks spread randomly?: A topological study of the global contagion network. *Plos one*, 15(9), e0238626.

Silva, J. S., and Tenreyro, S., 2006. The log of gravity. *The Review of Economics and Statistics*, 88(4), 641-658.

Seo, M. H., and Shin, Y. (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics*, 195(2), 169-186.

Seo, M. H., Kim, S., and Kim, Y. J. (2019). Estimation of dynamic panel threshold model using Stata. *The Stata Journal*, 19(3), 685-697.

Tinbergen,J., 1962. *Shaping the workd economy. Suggestions for an International Economic Policy.* New York: The Twentieth Century Fund.

Tong, H., 1990. *Non-linear time series: a dynamical system approach.* Oxford university press.

UNCTAD, 2022. *Global Trade Update, February 2022.*

UNWTO, H., 2020. *World Tourism Barometer and Statistical Annex, January 2020.*

5.7 — Appendix A

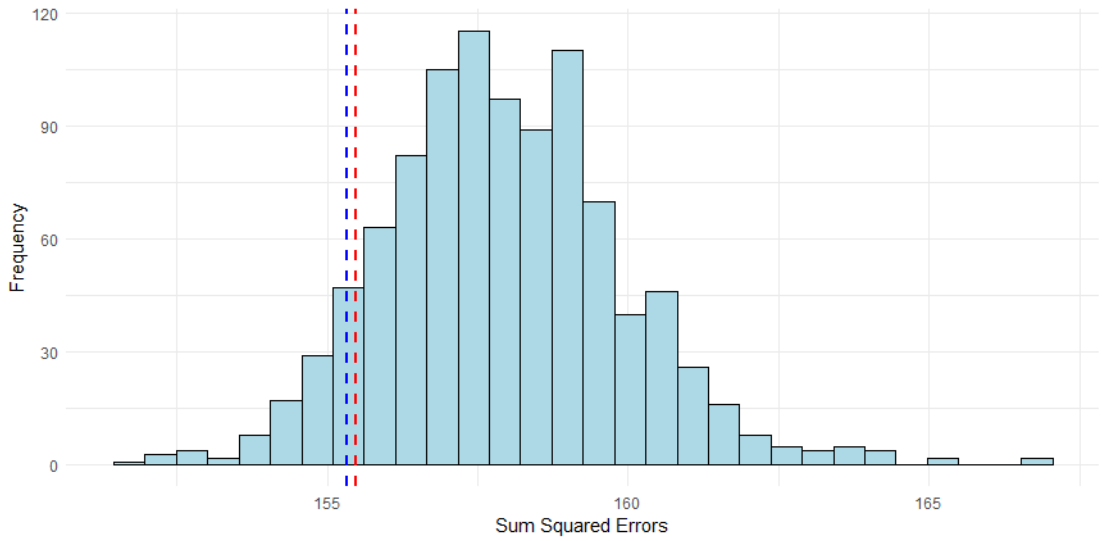
One of the main concerns in the estimation of VAR models is the rapid growth of parameters to be estimated as the number of equations increases, which can lead to a lack of degrees of freedom when estimating the model. In order to check that the configuration obtained in our estimation is credible, we examine the error generated by the Generalized Network Autoregressive (GNAR) model, when it was was trained on the specific obtained configuration, using data from the last year. We then compare this error with the errors generated by 1000 randomly generated configurations.

In a nutshell, the GNAR model assumes for a given network configuration, that each node of the network, i.e. each time series of industrial production of each country in our case, is driven by its own autoregressive process plus the influence of neighbors (i.e. other countries) up to r steps (this is the number of consecutive connections to arrive in the network from one node to other, or equivalently, from one country to another) modulated by the connection as a weight. Formally, the GNAR of order (p,s) $\in NxN_0^p$ for a $N \times 1$ vector of industrial production time series $Y_t = (Y_{1,t}, ..., Y_{N,t})^T$, based on the connection weights $C_{j \leftarrow i}$ between the series representing each node $i, j \in 1, ..., N$ (with $i \neq j$) and time $t \in 1, ..., T$, is characterized (assuming that no external covariate is part of the network) as

$$Y_{i,t} = \sum_{j=1}^{p} (\alpha_{i,j} Y_{i,t-j} + \sum_{r=1}^{s_j} \beta_{j,r} \sum_{j \in N_t^{(r)}(i)} C_{j \leftarrow i}^{(t)} Y_{q,t-j}) + u_{i,t}, \tag{C.14}$$

where $p \in \mathrm{N}$ is the maximum time lag, with $\mathbf{s} = (s_1, ..., s_p)$ where $s_j \in N_0$ is the maximum stage of neighbor dependence for time lag j and $N_0 = N \cup 0$, where $N_t^{(r)}(i)$ is the rth stage neighbor set of node i at time t, and where $w_{i,j}^{(t)}[0,1]$ is the connection weight between node i and node j at time t. Besides, $\beta_{j,r} \in R$ is the effect of the rth stage neighbors at lag j and $\alpha_{i,j} \in R$ are the autoregressive parameters at lag j for node i. The errors $u_{i,t}$ are assumed to be i.i.d. By calculating the error of the optimal model with the network configuration obtained by VAR modeling in the main configuration for a forecast horizon of two (in this case, it was a GNAR(2,2) model which minimizes the error, this is two autoregressive terms and influence of neighbors up to two steps of distance) when predicting for the years 2018 and 2019 two months ahead, and comparing it with the errors produced by 1000 randomly created configurations, Figure A1 shows that the error was found to be within the first decile of configurations with the lowest cumulative error. This is an indicator that the chosen configuration is correct, i.e. that the VAR estimation in the main manuscript is reliable.

Figure A1. Simulation of GNAR models. Prediction error distribution.



Notes. Blue dashed line represents the error of the GNAR(2,2) obtained for the VAR modeling configuration. Red dashed line represents first decile level.

5.8 — Appendix B

A comparative exercise of the different communities detected for the topologies obtained from three sectors with certain degree of openess, such as production, tourism and exports was performed with data from yearly growth of tourism and exports obtained from Eurostat. The available coincident sample for the three sectors spans from 1995 up to nowadays. As mentioned in the main text, to perform the analysis we filtered the edges with a higher level than 5 percent of connectivity for a node, as significant connections. From Table B1 it is clear that production and tourism topologies are more similiar in most of the descriptive features but in reciprocity, where tourism and exports present both a higher level. The former are more dense networks with larger interactions, while the network of exports is a less interactive network. Moreover, production is less prone to reciprocity in the interactions than tourism and exports.

Table B1. Network layers features.

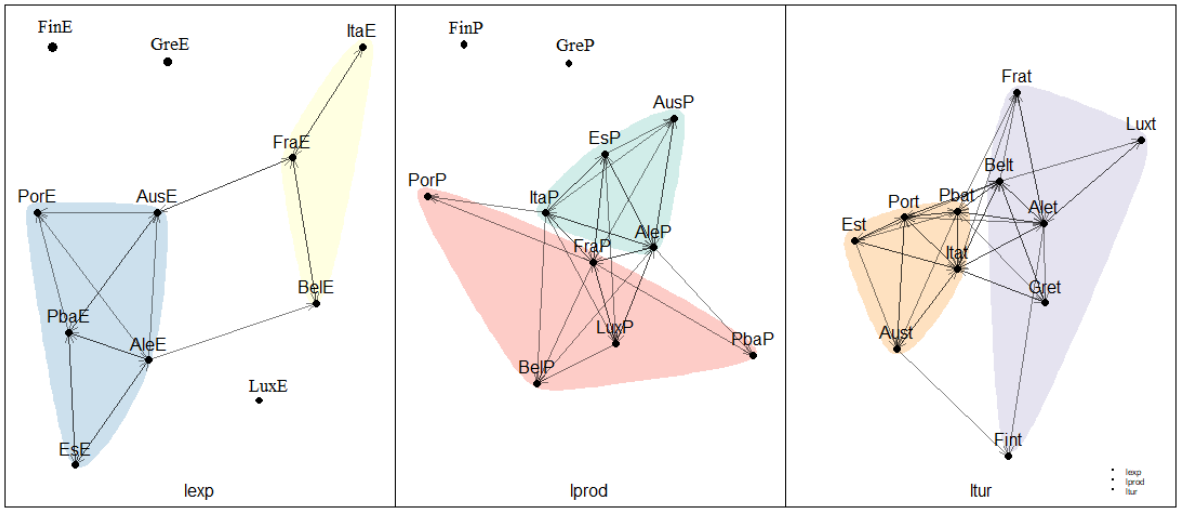
Layer	dens	cc	apl	diam	d	cen	re
Prod	0.43	0.69	0.19	0.45	0.08	0.36	0.58
Tur	0.43	0.65	0.20	0.40	0.06	0.36	0.72
Exp	0.20	0.50	0.30	0.61	0.13	0.17	0.74

Notes. *dens* makes reference to the density of edges in the network, *cc* to the clustering coefficient, *apl* to the average path length, *diam* to the diameter, $\bar{d}$ to the mean distance between nodes, *cen* to centrality and *re* to reciprocity.

In this vein, and based on the histogram of the degree of the nodes (i.e. countries), a measure of dissimilarity can be obtained through analyzing the divergence of the distributions, for instance the Jeffrey dissimilarity degree measure (Jeffrey, 1948). In particular, and limited from 0 (in case of overlapping) to 1 (in case of total dissimilarity), the networks of production and tourism presented a low dissimilariy of 0.03, while production and exports, as tourism and exports, presented a higher level, 0.42 and 0.46, respectively, as expected.

On the other hand, Figure B1 depicts how the Louvain partitioning algorithm shows different communities in the different sectors. While in general more distant countries such as Greece and Finland are more isolated, in tourism, for instance, mediterranean countries such as Italy, Portugal or Spain share a community, while the three of them do not share community in exports or production. The two main exporter countries, as The Netherlands and Germany are, share community in the exports network. On the other hand, it must be noted that France and Belgium are the only countries that share community in all three sectors, indicating a constant relationship in all three analyzed economic areas.

Figure B1. Community detection through Louvain algorithm.



Notes. The figure on the left represents the communities for the exports network, while the figure on the center represents the communities for the production, and the figure of the right to the tourism sector.

5.9 — Appendix C

When analyzing the different spillovers, two qualitative observations are that closer countries seems to have a greater influence, as well as that, apparently, larger countries exert a greater influence on smaller countries than vice versa. As a consequence, a simple first approach would to run a static gravity model with the variables GDP and distance,

$$C_{j\leftarrow i} = \beta_0 GDP_i^{\beta_1} GDP_j^{\beta_2} Dist_{j\leftarrow i}^{\beta_3} u_{j\leftarrow i}, \tag{C.15}$$

where $Dist_{j\leftarrow i}$ is measured as the distance between the capital in country i to the closer border of country j, GDP represents the size of an economy, and $C_{j\leftarrow i}$ the spillover from country i to country j. Thus, we might fit the model using PPML estimation whose coefficients can be interpreted as elasticities. However, it might be the case that estimations can be biased due to hidden factors that are not taken into account, such as common multinationals, trade partnerships, or industrial composition. Within this rationale, Chen, Fernández-Val, and Weidner (2021) propose a nonlinear panel models with factor structures in the unobservables. In particular, they apply their technique to the gravity model with an estimate based on a Poisson maximum likelihood estimate and allows to infer the presence of latent variables not being taken into account. Within our case, this can be written as

$$E[C_{j\leftarrow i}|GDP_i, GDP_j, Dist_{j\leftarrow i}, \beta_0, \alpha_{1i}, \gamma_{1i}] = exp(\beta_0 + \beta_1 GDP_i + \beta_2 GDP_j + \beta_3 Dist_{j\leftarrow i} + \alpha_{1i}'\gamma_{1j}), \tag{C.16}$$

where α_{1i} and γ_{2i} are vectors of factors and factor loadings, and where up to R factors can be included. If there are no variables without taking into account the estimates with or without factors will be similar. Nevertheless, if the estimates differ, the presence of unaccounted-for factors is confirmed and the model in equation C.15 can be extended to include more variables.

Therefore, to verify whether we might be leaving variables out of the analysis when just considering GDP and distance, we fitted the nonlinear panel model with factor structures and unobservable effects. Due to the data availabliity of others variables that will be used later, we fitted the model with the sample which already covers all countries within the EU, from 1995 and up to 2019 to avoid Covid-19 influence, although the results are qualitative consistent with the whole sample [3]. As can be noted from the first three columns of Table C1, the fitted Poisson model with unobserved structures confirms the expected signs of the coefficients, even unveiling a negative sign for the GDP of the receiver country, and more importantly, through the inclusion of one and two factors, the slightly different estimations point to some missing covariates in the model.

[3]We translated the original code from the authors of the model in Matlab to R, available at Github.

Table C1. Estimation Static Gravity model with unobserved structures.

	2000-2019			1970-2019
	R=0	R=1	R=2	R=0
$log(dist_{ij})$	-0.23 [0.14]	-0.28 [0.15]	-0.33 [0.17]	-0.31 [0.22]
$log(GDP_i)$	1.41 [6.89]	0.99 [7.21]	0.33 [8.78]	0.53 [21.76]
$log(GDP_j)$	-1.31 [7.01]	-1.19 [7.33]	-0.76 [8.98]	-4.82 [17.22]
Log-likelihood	-0.49	-0.46	-0.44	-0.68
Beta precision	10^{-7}	10^{-3}	10^{-3}	10^{-9}
Max reps	26	6	5	9
Max steps	250	50	50	75

Notes. *dist* makes reference to physical distance between the capital of country i to the closer border of country j, *GDP* refers to Gross Domestic Product, and R makes reference to the number of hidden factor structures. Beta precision, Max reps and Max steps make reference to the convergence criteria in the estimation. Number in brackets refer to standard deviations of the coefficients.

5.10 — Tables

Table 1. EMU business cycles connectivity matrix: 1966M1-2022M11. 2 months ahead forecast

	Austria	Belgium	Finland	France	Germany	Greece	Italy	Luxembourg	The Netherlands	Portugal	Spain	From others
Austria	5.29	0.26	0.10	0.47	1.13	0.02	0.75	0.34	0.17	0.06	0.50	3.80
Belgium	0.38	5.30	0.08	0.80	0.81	0.02	0.66	0.64	0.08	0.01	0.31	3.80
Finland	0.16	0.10	7.70	0.04	0.20	0.04	0.31	0.16	0.06	0.11	0.20	1.39
France	0.34	0.40	0.01	4.25	0.87	0.06	1.26	0.57	0.16	0.15	1.03	1.39
Germany	0.50	0.44	0.08	1.07	4.60	0.01	0.83	0.47	0.19	0.27	0.65	4.49
Greece	0.04	0.01	0.02	0.16	0.16	7.76	0.22	0.01	0.11	0.01	0.59	1.33
Italy	0.42	0.29	0.09	1.06	0.59	0.14	4.90	0.42	0.15	0.26	0.77	4.19
Luxembourg	0.37	0.45	0.06	0.72	0.68	0.01	0.55	5.45	0.11	0.20	0.50	3.64
The Netherlands	0.29	0.11	0.03	0.48	0.61	0.01	0.43	0.20	6.42	0.06	0.45	2.67
Portugal	0.09	0.05	0.09	0.56	0.41	0.01	0.51	0.16	0.04	6.8	0.42	2.33
Spain	0.23	0.14	0.02	1.15	0.62	0.09	1.08	0.37	0.11	0.39	4.89	4.20
To others	2.82	2.26	0.57	6.52	6.07	0.39	6.60	3.33	1.18	1.52	5.42	36.69
Total	8.11	7.55	8.28	10.76	10.67	8.15	11.50	8.78	7.60	8.28	10.31	100

Notes. Each element $D_{ij}^{h} = \frac{d_{ij}^{h}}{N}$ represents the percentage of the total forecast error variance.

Table 2. EMU business cycles connectivity matrix: 1966M1-2022M11. 12 months ahead forecast

	Austria	Belgium	Finland	France	Germany	Greece	Italy	Luxembourg	The Netherlands	Portugal	Spain	From others
Austria	2.51	0.40	0.23	0.73	1.51	0.21	0.99	0.80	0.54	0.13	1.03	6.58
Belgium	0.30	3.27	0.16	0.96	1.33	0.09	0.94	0.87	0.37	0.05	0.77	5.82
Finland	0.18	0.22	4.02	0.44	0.37	0.22	0.75	0.84	0.64	0.16	1.24	5.07
France	0.30	0.27	0.02	2.65	1.14	0.33	1.31	0.81	0.49	0.29	1.49	6.44
Germany	0.36	0.39	0.07	1.02	3.77	0.04	0.85	0.83	0.55	0.28	0.93	5.32
Greece	0.10	0.03	0.03	0.25	0.14	5.56	0.36	0.09	0.48	0.25	1.81	3.53
Italy	0.20	0.24	0.13	1.12	0.64	0.54	3.76	0.76	0.36	0.20	1.13	5.33
Luxembourg	0.23	0.53	0.03	1.00	0.86	0.04	0.59	4.42	0.31	0.21	0.87	4.67
The Netherlands	0.21	0.20	0.05	0.62	0.93	0.24	0.75	0.59	4.23	0.11	1.16	4.86
Portugal	0.13	0.04	0.05	1.01	0.47	0.43	0.68	0.45	0.08	4.79	0.96	4.30
Spain	0.21	0.09	0.04	1.13	0.57	0.45	1.09	0.55	0.27	0.53	4.16	4.93
To others	2.20	2.41	0.82	8.28	7.96	2.58	8.32	6.59	4.10	2.21	11.38	56.85
Total	4.71	5.68	4.83	10.93	11.73	8.14	12.08	11.01	8.33	6.99	15.55	100

Notes. See Table 1.

Table 3. EMU business cycles Net Connectedness: 1966M1-2022M11. 2 months ahead forecast

	To others	From others	Net Connectedness
Austria	2.82	3.80	-0.98
Belgium	2.26	3.80	-1.54
Finland	0.57	1.39	-0.81
France	6.52	4.85	-1.67
Germany	6.07	4.49	1.58
Greece	0.39	1.33	-0.94
Italy	6.60	4.19	2.41
Luxembourg	3.33	3.64	-0.31
The Netherlands	1.18	2.67	-1.49
Portugal	1.52	2.33	-0.81
Spain	5.42	4.20	1.22

Notes. Countries with negative (positive) net connectivity are business cycles receivers (transmitters).

Table 4. EMU business cycles Net Connectedness: 1966M1-2022M11. 12 months ahead forecast

	To others	From others	Net Connectedness
Austria	2.20	6.58	-4.38
Belgium	2.41	5.82	-3.41
Finland	0.82	5.07	-4.26
France	8.28	6.44	1.83
Germany	7.96	5.32	2.64
Greece	2.58	3.53	-0.95
Italy	8.32	5.33	2.99
Luxembourg	6.59	4.67	1.92
The Netherlands	4.10	4.86	-0.75
Portugal	2.21	4.30	-2.09
Spain	11.38	4.93	6.46

Notes. See Table 3.

Table 5. Differences on EMU business cycles connectivity matrix from sample up to 2022M11 to 2019M12. 2 months ahead forecast

	Austria	Belgium	Finland	France	Germany	Greece	Italy	Luxembourg	The Netherlands	Portugal	Spain	From others
Austria	-1.29	0.04	-0.03	0.17	0.14	0.00	0.51	0.12	-0.00	0.05	0.29	1.29
Belgium	0.07	-0.56	-0.03	0.07	0.01	-0.01	0.22	0.13	0.02	0.01	0.06	0.56
Finland	0.04	0.01	-0.06	-0.02	-0.04	0.00	0.02	0.01	0.01	0.00	0.02	0.06
France	0.13	-0.09	-0.04	-1.12	0.35	-0.03	0.49	0.09	-0.06	0.08	0.19	1.12
Germany	0.19	-0.11	-0.09	0.46	-1.45	0.00	0.52	0.12	-0.04	0.11	0.31	1.45
Greece	0.02	0.00	-0.00	0.09	0.07	-0.37	0.08	0.01	-0.01	0.01	0.10	0.37
Italy	0.25	0.03	-0.09	0.19	0.22	-0.05	-1.10	0.15	-0.01	0.11	0.29	1.10
Luxembourg	0.12	0.03	-0.02	0.17	0.12	0.00	0.31	-0.91	0.02	0.03	0.13	0.91
The Netherlands	0.07	0.01	0.00	-0.02	0.03	-0.01	0.12	0.05	-0.32	0.01	0.04	0.32
Portugal	0.08	0.03	-0.03	0.37	0.25	-0.00	0.38	0.05	-0.01	-1.41	0.29	1.41
Spain	0.20	0.02	-0.01	0.34	0.34	-0.01	0.60	0.12	-0.04	0.18	-1.74	1.74
To others	1.16	-0.03	-0.33	1.81	1.51	-0.10	3.26	0.86	-0.12	0.58	1.72	10.34
Total	-0.13	-0.59	-0.39	0.70	0.06	-0.46	2.16	-0.05	-0.43	-0.83	-0.02	0

Table 6. EMU business cycles Dynamic Connectedness: 1966M1-2022M11 for different rolling windows

	$w = 48$	$w = 54$	$w = 60$	$w = 66$	$w = 72$	$w = 78$	$w = 84$	$w = 90$	$w = 96$
Mean Connectedness Index	56.46	53.17	50.51	48.37	46.63	45.19	43.97	42.96	42.13
Mean Difference during Expansions	-18.26	-21.15	-5.53	-5.66	-2.60	5.31	5.47	-4.09	-1.22
Mean Difference during Recessions	46.95	54.15	41.68	40.67	36.41	30.08	32.29	35.82	40.66

Table 7. Panel Unit Root tests.

	Levinlin.	IPS	Madwu	Hadri
$Spillover_{ij}$	-21.10 [$< 10^{-2}$]	-26.97 [$< 10^{-2}$]	1245.70 [$< 10^{-2}$]	483.28 [$< 10^{-2}$]
GDP_i	16.35 [1.00]	15.69 [1.00]	37.45 [1.00]	1742.3 [$< 10^{-2}$]
Exp_{ij}	-7.61 [$< 10^{-2}$]	-19.89 [$< 10^{-2}$]	1215.1 [$< 10^{-2}$]	337.67 [$< 10^{-2}$]
Tur_{ij}	-6.36 [$< 10^{-2}$]	-9.16 [$< 10^{-2}$]	732.67 [$< 10^{-2}$]	498.92 [$< 10^{-2}$]
Sim_{ij}	-14.78 [$< 10^{-2}$]	-20.54 [$< 10^{-2}$]	1121.6 [$< 10^{-2}$]	399.65 [$< 10^{-2}$]
$log(Spillover_{ij})$	-36.95 [$< 10^{-2}$]	-39.09 [$< 10^{-2}$]	2045.1 [$< 10^{-2}$]	384.00 [$< 10^{-2}$]
$log(GDP_i)$	1.67 [0.95]	-10.26 [1.00]	86.01 [1.00]	2178.2 [$< 10^{-2}$]
$log(Exp_{ij})$	-10.78 [$< 10^{-2}$]	-21.35 [$< 10^{-2}$]	1283.10 [$< 10^{-2}$]	357.62 [$< 10^{-2}$]
$log(Tur_{ij})$	-9.99 [$< 10^{-2}$]	-13.48 [$< 10^{-2}$]	750.03 [$< 10^{-2}$]	395.45 [$< 10^{-2}$]
$log(Sim_{ij})$	-14.47 [$< 10^{-2}$]	-20.39 [$< 10^{-2}$]	1114.80 [$< 10^{-2}$]	397.75 [$< 10^{-2}$]
$\Delta log(GDP_i)$	-23.20 [$< 10^{-2}$]	-50.73 [$< 10^{-2}$]	3132.00 [$< 10^{-2}$]	62.55 [$< 10^{-2}$]

Notes. Figures in brackets make reference to p-values. Every test includes intercept and trend. For Levinlin test the H_0 is non stationarity for all individuals and the distribution is t. For IPS the H_0 is non stationarity and the distribution is the modified t. For Maddala-Wu test the H_0 is non stationarity and the distribution is χ^2.For Hadri-LM test the H_0 is that any of the individuals has a unit root and the distribution is $N(0,1)$. Lags of ADF regressions within the tests are selected through AIC criteria with the maximum lag set equal to 2.

Table 8. Estimation of alternative explanatory panel models.

Parameter	Model 1	Model 2	Model 3 High	Model 3 Low
$\Delta log(GDP_i)$	0.38 (0.71)	-16.31 ($< 10^{-2}$)	-38.38 ($< 10^2$)	10.88 (0.38)
$\Delta log(GDP_j)$	-0.92 (0.37)	2.14 (0.14)	1.81 (0.40)	-9.49 (0.19)
$log(Exp_{ij})$	0.07 ($< 10^2$)	$3.10 \cdot 10^{-3}$(0.28)	1.18 ($< 10^2$)	-0.24 (0.03)
$log(Tur_{ij})$	0.04 ($< 10^2$)	0.04 ($< 10^2$)	-0.53 (0.02)	-0.45 ($< 10^2$)
$log(Sim_{ij})$	0.91 ($< 10^2$)	0.26 (0.33)	2.84 (0.41)	-3.63 (0.08)
$log(dist_{ij})$	-0.10 ($< 10^2$)	-0.14 ($< 10^2$)	-	-
Intercept	-1.79 ($< 10^2$)	-	-26.71 ($< 10^2$)	-
$log(Spill_{ij,t-1})$	-	0.53($< 10^2$)	-0.33 ($< 10^2$)	0.18 ($< 10^2$)
q_{it}	-	-	54.20 ($< 10^2$)	-

Notes. Model 1 refers to unbalanced gravity model with time dimension for period 1995-2022 estimated through PPML. Model 2 refers to unbalanced dynamic panel model for period 1995-2022 estimated through system GMM. Model 3 makes reference to balanced non linear TAR dynamic model for period 2000-2022. Numbers in parenthesis represents the p-value of the coefficients.

5.11 — Figures

Figure 5.1. Industrial similitude matrix .

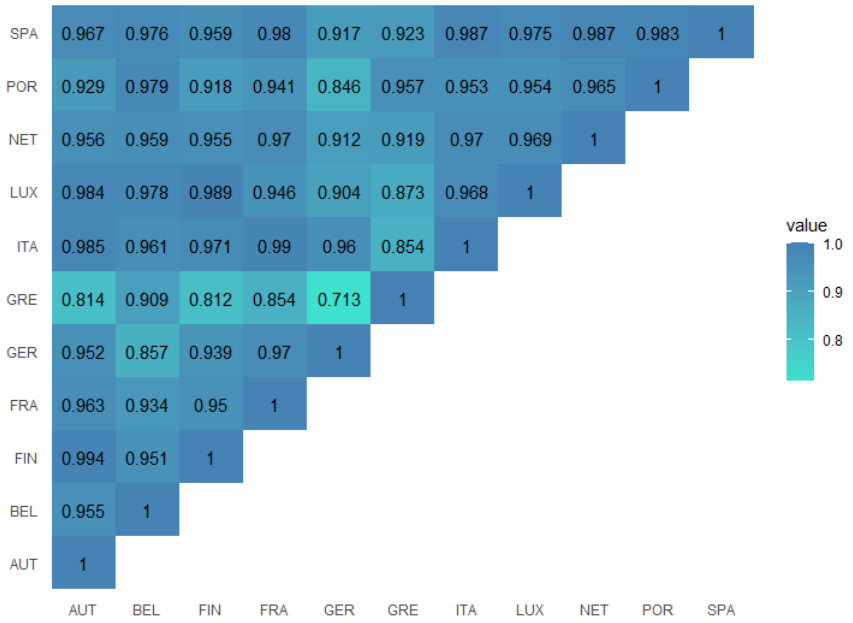
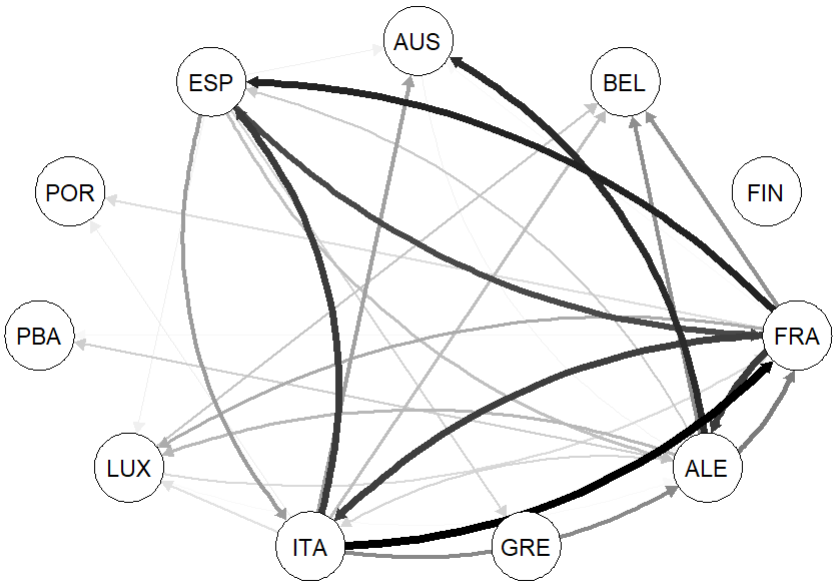


Figure 5.2. Static connectedness network between EMU countries at 2-months horizon.



Notes. Arrows indicate the sense and the magnitude of the shocks transmission. The darker the higher is the relative ability of a country to influence on the other. Arrows with a level below 0.05 are not represented.

Figure 5.3. Community detection in industrial production network for a 2-months horizon.

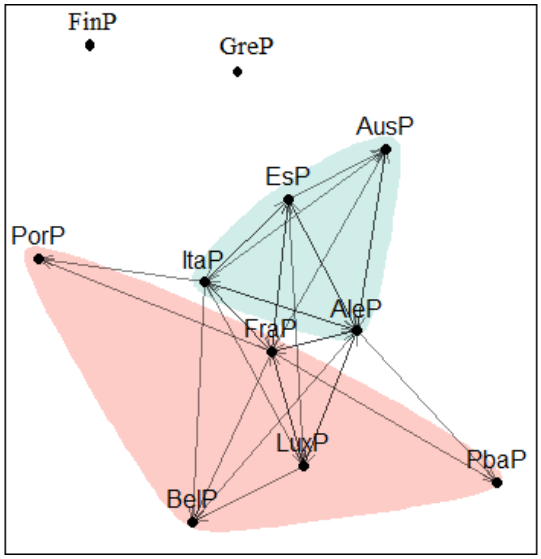


Figure 5.4. Total Connectedness for different horizons.

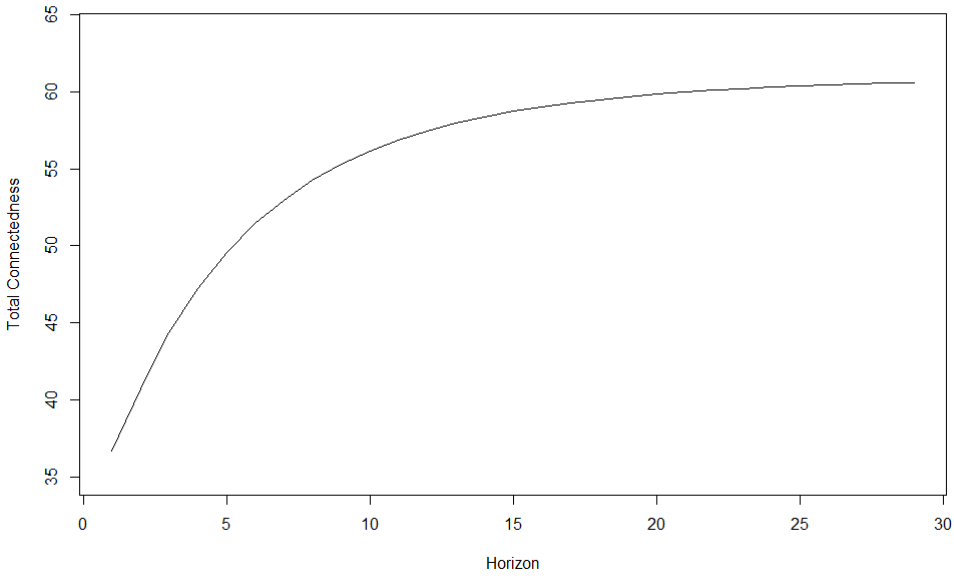
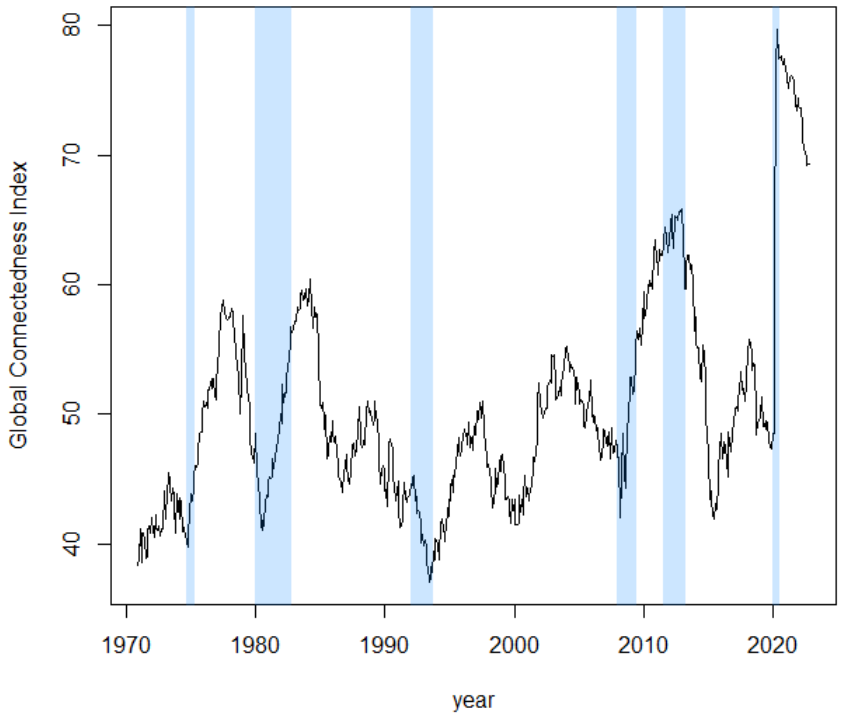
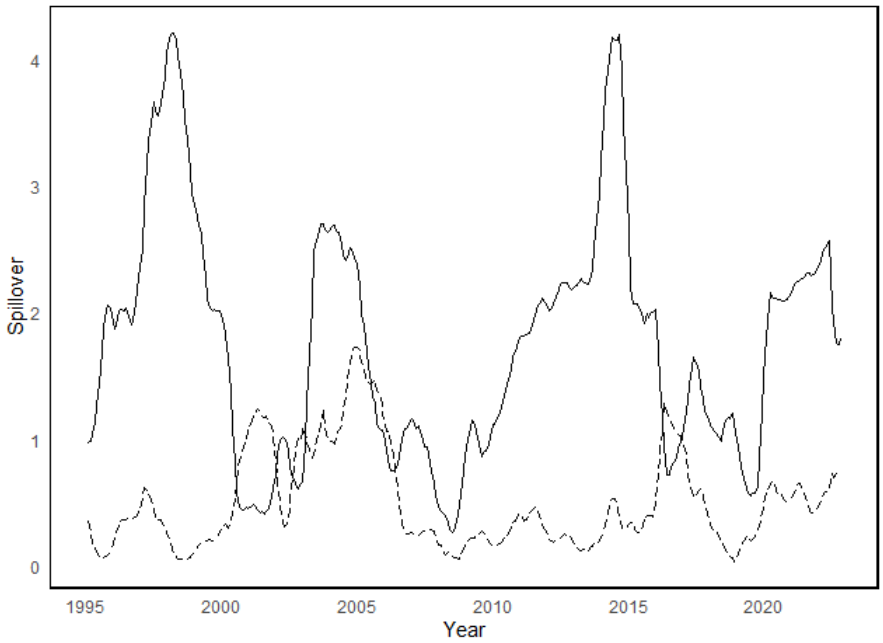


Figure 5.5. Dynamic Connectedness index EMU countries.



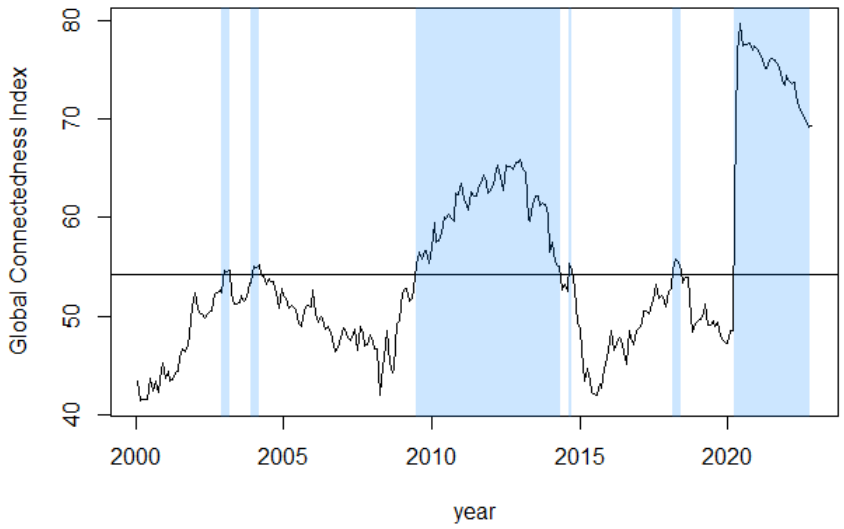
Notes. Window of $w = 60$ period and forecast horizon of $h = 2$ periods. Shaded areas represent periods of recession according to the Euro Area Business Cycle Network (EABCN) Comittee

Figure 5.6. Spillovers Germany to France and Greece to Germany 1995-2022.



Notes. Solid line represents the Spillover from Germany to France. Dashed line represents the Spillover from Greece to Germany

Figure 5.7. Global Connectedness Threshold and regions



Notes. Shadow areas represent the regions inferred from TAR model.

Conclusions

The greater availability of data has generated an increasing interest in describing economic problems, and in particular, in establishing econometric models. Although the linear approximation is useful for simplifying problems and analyzing behavior from a descriptive perspective, in many cases the relationships, or responses, are not strictly linear, with the respective loss of precision. The economy, as a complex system with multiple interdependencies, is a discipline conducive to more complex modeling to capture these relationships.

This dissertation illustrates the benefit of using nonlinear techniques for the description of certain problems, with special emphasis on time series description. Its use is made in different contexts, and the application of nonlinearities is adapted to each problem of the data used. In general, their use results in a benefit in terms of accuracy and analytical capability, which is illustrated by comparative analysis with other techniques, simulations, and accuracy measurements.

In the second chapter, it is shown how the adoption of nonlinear modeling is useful in cases of a data generating process with influenceable observations. Specifically, a nonlinear and nonparametric approach was proposed to deal with the problem of certain influential observations when fitting a model in a univariate context. In particular, the problem was focused on the prediction of recessions after the one that occurred, of exceptional magnitude, with the Covid-19 pandemic.

The problem of its presence in linear models was illustrated, and consequently, a nonlinear and nonparametric extension of Wecker's linear approximation (1979) was proposed, which also includes a weighting system based on symbolic analysis to make predictions. By means of simulation and comparison, it is shown and quantified that the proposal is more robust than the linear alternatives, and that although before the Covid-19 influential observation its performance was similar to that of the linear models, by incorporating this observation its performance is more advisable. The approach gives rise to work on a future generalization to a multivariate context.

In the third chapter, we show how the use of nonlinear techniques is useful to analyze more complex databases with interrelationships of different natures. In particular, we adopt the use of the boosting-based decision trees technique by Friedman (2002) to analyze economic recessions in Spain, but in this case by using a large and more complex database. This is a particularly relevant analysis due to the extraordinary economic consequences that both

the Great Recession and the Covid-19 pandemic had on the Spanish economy. Through this technique, more complex relationships between variables can be unveiled, further than those that are solely linear.

By applying this technique to a database of 270 indicators, the fit of the recession prediction for a horizon of three and six month ahead turns out to be accurate. In addition, the use of this technique has provided us with a tool to evaluate which indicators are more relevant at each moment to make the prediction, being able to analyze in cases of correct forecasts which indicators anticipate recessions, and also to graphically analyze interaction effects between variables against the variable to be predicted.

Throughout the fourth chapter, it becomes clear how the inclusion of certain nonlinearities can help in the description of time series dynamics. In particular, the proposed nonlinear Kalman filter of Brockwell and Davis (2009) is applied to dynamic factor models to the nowcasting problem of homicides with firearm, so that forward information from different unofficial databases can be exploited to describe the official series, which has a publication lag of up to 23 months.

After a model fitting exercise to determine the variables useful to describe the dynamics of gun homicides within a pool of collected variables, through a pseudo-real time exercise, the fitted model was compared with different linear modeling, as well as with other additional nonlinear approaches, which turned unable to have a good description due to the paucity of the database. In this sense, the best predictive capacity of the proposed approach was statistically determined, demonstrating that a certain introduction of nonlinearity, adjusted to the database in question, can be beneficial. The tool can be of great utility for tracking homicides, both for lawmakers and for economic stakeholders with a special interest in it, such as insurers or real estate agents. Of course, the approach has limitations, and further analysis is needed to explore specifications with a higher degree of non-linearity.

Finally, the fifth chapter shows the capacity of a nonlinear model in the description of a problem of great interest such as that of connectivities between countries. After determining the spillovers between the industrial productions of eleven EMU countries by means of a linear approximation, the persistence and asymmetry of the behavior of these relationships based on the economic cycle provide a favorable framework for the adjustment of a non-linear model.

In particular, in order to be able to describe this asymmetry, a TAR model (1978) describing two different regimes was fitted based on a variable generated by inferring the spillovers between countries, the global connectivity index variable of the countries in the sample. We saw the importance of variables such as tourism, and we also saw how others such as exports or GDP played a different role depending on the regime in which the dynamics are found. The approximation was corroborated by tests of linearity and exogeneity of instruments in the estimation. This approach also allowed us to create a connectivity indicator, which has

implications for effective decision-making at certain points in the cycle.