Journal of International Money and Finance 133 (2023) 102824

Contents lists available at ScienceDirect

Journal of International Money and Finance

journal homepage: www.elsevier.com/locate/jimf

Macro-financial spillovers ☆

John Cotter [a], Mark Hallam [b,*], Kamil Yilmaz [c]

[a] Graduate School of Business, University College Dublin, Ireland
[b] Department of Economics and Related Studies, University of York, UK
[c] Department of Economics, Koç University, Turkey

ARTICLE INFO

Article history:
Available online 6 March 2023

JEL codes:
G10
G01
E30
E44
C13
C32

Keywords:
Spillovers
Connectedness
Macro-financial
Mixed-frequency
Forecasting

ABSTRACT

We analyse spillovers between the real and financial sides of the US economy, and between those in the US and other advanced economies. The approach developed allows for differences in sampling frequency between financial and macroeconomic data. We find that financial markets are typically net transmitters of shocks to the real side of the economy, particularly during turbulent market conditions. This result holds both for domestic US macro-financial spillovers, and also those between the US and other advanced economies. Our macro-financial spillover measures are found to have significant predictive ability for future macroeconomic conditions in both in-sample and out-of-sample forecasting environments. Furthermore, the predictive ability frequently of our macro-financial measures frequently exceeds that of purely financial systemic risk measures previously employed in the literature for the same task.

1. Introduction

Over the last decade, there has been significant interest in financial contagion, spillovers and systemic risk, both in the academic community and monetary and financial regulatory authorities. The 2008–2009 financial crisis was previously a key driver of this interest, and the current COVID-19 pandemic has also highlighted its importance. A primary topic of research within this literature has been the development of quantitative measures and statistical tests for spillovers and systemic risk in financial markets, with examples including Allen et al. (2012),Billio et al. (2012),Diebold and Yılmaz (2014), Adams et al. (2014),Engle et al. (2015), Adrian and Brunnermeier (2016) and Brownlees and Engle (2017).

☆ The authors wish to thank participants of the 2016 FMA Annual Meeting in Las Vegas, the 2016 Financial Econometrics and Empirical Asset Pricing Conference at Lancaster University, the Barcelona GSE Summer Forum at Universitat Pompeu Fabra, the 11th Computational and Financial Econometrics Conference in London, the 5th Asset Pricing Workshop at the University of York, the 2018 Irish Academy of Finance conference, the 2020 FMA Online Annual Meeting, seminar participants at Bilkent University, Boğaziçi University, University College Dublin, University of Essex, University of Konstanz, University of Liverpool, University of Stirling, the Central Bank of Ireland, University of Groningen, Aarhus University, European Securities and Markets Authority, Vienna University of Economics and Business, and Abhinav Annand, Turan Bali, Michael Brennan, Christian Brownlees, Eric Ghysels, Maureen O Hara, Matt Spiegel and Josef Zechner for valuable feedback and comments. John Cotter and Mark Hallam gratefully acknowledge the support of Science Foundation Ireland under grant number 16/SPP/3347 and 17/SPP/5447. Kamil Yilmaz and Mark Hallam gratefully acknowledge the support of The Scientific and Technological Research Council of Turkey under grant number TUBITAK 114K954.

* Corresponding author at: Department of Economics and Related Studies, University of York, York, England, UK.
E-mail addresses: john.cotter@ucd.ie (J. Cotter), mark.hallam@york.ac.uk (M. Hallam), kyilmaz@ku.edu.tr (K. Yilmaz).

https://doi.org/10.1016/j.jimonfin.2023.102824
0261-5606/

This literature on the quantitative measurement of spillovers and systemic risk has provided valuable insights into the strength and structure of financial market linkages at both the firm and market-level, but has largely ignored the real or non-financial side of the economy. This leaves unanswered the key question of the magnitude and direction of the linkages between Main Street (the real economy) and Wall Street (financial markets) and how they vary over time. This paper directly estimates spillovers between the financial and real sides of the economy with a new approach designed specifically for this macro-financial context.

We perform a detailed empirical investigation of macro-financial spillovers for the US economy, estimating both the aggregate level of US macro-financial spillovers and also disaggregated directional measures in a dynamic estimation environment. This allows us to decompose spillovers in detail and analyse how the level and composition of macro-financial spillovers have evolved over time. We then expand this analysis to consider pairs of economies comprised of the US and one of five other advanced economies, allowing us to study the international dimensions of macro-financial spillovers. Finally, we also investigate the predictive ability of our macro-financial spillover measures for future US macroeconomic conditions, finding that they outperform existing purely financial systemic risk measures used for this task in the literature.

Despite the previously noted lack of work on the quantitative measurement of macro-financial spillovers specifically, there is an extensive literature on macro-financial linkages in a more general sense that helps to motivate the current work. Examples include, amongst others, Ellington (2018) who examines the links between financial market illiquidity and macroeconomic dynamics for UK data; Prieto et al. (2016) who examine time variation in linkages between US GDP and financial data including stock prices and credit spreads; Galvão and Owyang (2018) who detail the links between a panel of key financial variables and industrial production growth and inflation; and Hubrich and Tetlow (2015) who examine the behaviour of the macroeconomy under periods of varying financial stress.

A second, and generally more structural, strand of the literature focuses on specific examples of macro-financial linkages. For example, a deterioration in financial conditions is expected to negatively impact the real side of the economy through a reduction in the willingness of financial firms to extend credit to corporate clients, which may, in turn, suppresses investment and thus economic output. This is supported empirically by work such as Ivashina and Scharfstein (2010),Cingano et al. (2016) and Li et al. (2019). Likewise, shocks to the real side of the economy may feed back into financial markets by increasing corporate defaults or reducing firm equity values. In particular, there is significant literature studying the response of financial markets to macroeconomic news and announcements, such as the work of Brenner et al. (2009) and Savor and Wilson (2013). Furthermore, adverse feedback loops may develop between the financial and real sides of the economy, amplifying the effect of negative shocks through mechanisms such as the financial accelerator of Bernanke et al. (1996). Given that we seek to estimate and analyse macro-financial shock transmission and spillovers in a general sense arising as a result of all these linkages collectively, our work lies closer to the non-structural side of the literature noted in the previous paragraph than these more structural studies.

Our work also clearly has strong links to the literature on the estimation and quantitative measurement of systemic risk and spillovers in financial markets such as Allen et al. (2012), and others cited above, but in a macro-financial context. The vast majority of the past work in this field has ignored the non-financial side of the economy, despite the potential importance of the macro-financial dimension being frequently noted (see, for example, Brunnermeier et al., 2011). The few exceptions include Baur (2012) and Claessens et al. (2012); however, in these cases, the real side of the economy enters only via financial data for non-financial firms, rather than through conventional aggregate real economy series. Similarly, both Allen et al. (2012) and Brownlees and Engle (2017) create non-financial variants of their respective systemic risk indices by utilising data for non-financial firms. However, these non-financial indexes are used for robustness checks and minor extensions to the core empirical analysis, not to study the interaction between the two sides of the economy as we do here. The disadvantage of such approaches in our context is that financial data for non-financial firms provide a less direct and much narrower measure of conditions in the real sector of the economy than aggregate macroeconomic series.

Our approach is thus explicitly developed for combinations of financial and real economy[1] variables. A key practical issue encountered in this context is that most financial series are available at much higher sampling frequencies than the monthly or quarterly frequencies of most real economy series. Standard econometric methods for the estimation of models in such a context impose the use of a common frequency for all indicators, requiring the financial series to be aggregated to the lower frequency of the real economy series, discarding potentially relevant high-frequency information. Our approach instead uses mixed-frequency methods developed by Ghysels (2016) to allow monthly macroeconomic series to be used alongside weekly financial series. Ghysels (2016) and Schorfheide and Song (2015), amongst others, have found that in several settings, mixed-frequency approaches can provide gains in estimation and forecasting compared to a common-frequency approach.

For the quantitative measurement of spillovers, we employ the Diebold-Yilmaz spillover measures previously used in the context of financial markets (see Diebold and Yılmaz 2009,2012 and 2014), but we modify and extend the approach appropriately for the mixed-frequency context we work in. The Diebold-Yilmaz (DY) approach provides a set of spillover measures at various levels of aggregation. The most aggregated measure is the total spillover index, which provides a single numerical value that measures the overall level of spillovers between the series of interest and is thus analogous to many of the other financial spillover or systemic risk measures proposed in the literature, such as the CATFIN and SRISK measures of Allen et al.

[1] Throughout the current work, we use the terms 'real economy' and 'macroeconomic' interchangeably when referring specifically to series and variables. More generally, the term macroeconomic could also refer to certain financial series types, but this is not the case in the current context.

(2012) and Brownlees and Engle (2017) respectively. We also employ the most disaggregated pairwise spillover indexes, which provide a set of indexes measuring spillovers from a specific series to each of the other series. The pairwise measures are directional in nature and so allow for a detailed analysis of spillover structure that would not be possible with non-directional alternative measures that account for the majority of the existing literature. This ability to construct a set of pairwise directional spillover indexes is also shown to be valuable when using our measures to forecast future macroeconomic conditions.

In our analysis of US macro-financial spillovers we focus on equity and bond markets for the financial side of the economy and a broad measure of economic conditions for the real side of the economy, namely the Chicago Fed National Activity Index (CFNAI). Our sample period spans 1980–2020, thus including many significant economic and financial events such as the 2008–2009 financial crisis and the beginning of the COVID-19 pandemic. The dynamics and magnitude of estimated US macro-financial spillovers obtained from our new mixed-frequency extension of the DY approach differ significantly from those obtained from a more straightforward, but otherwise equivalent, common-frequency modelling approach that discards additional high-frequency financial information. Perhaps most notably, the magnitude of spillovers estimated by our new mixed-frequency DY approach is typically substantially higher than that implied by the similar common-frequency approach. This finding suggests that using a common-frequency modelling approach that discards additional high-frequency financial information within each month results in estimated macro-financial connectedness being lower on average. Exploiting this additional financial information also results in estimated spillover indexes whose dynamics are generally more consistent with known historical events, such as those during the 2008–2009 financial crisis, than common-frequency estimates.

We also find that financial markets are typically net transmitters of shocks to the real economy regardless of whether our mixed-frequency approach or the equivalent common-frequency approach is used, but that the role of financial markets in spillover transmission is more significant when employing the former. These findings are particularly evident when markets face turbulence, most notably during the 2008–2009 crisis and the COVID-19 pandemic.

Following the US analysis, we study international macro-financial spillovers between the US and other major advanced economies. For this component of the analysis we study pairs of countries consisting of the US and one of five other advanced economies, namely Canada, France, Germany, Japan and the UK. In each case we employ returns on a major national equity index for each of the countries and industrial production growth for each country for the financial and non-financial sides of the economy respectively. The key findings for the US are also shown to hold more generally in this international context. Firstly, estimates of the overall level of international macro-financial spillovers are again substantially higher when employing our mixed-frequency approach than the equivalent common-frequency approach. Second, cross-country spillovers from financial variables constitute the majority of overall spillovers on average over the sample period. However, we find that cross-country real economy to real economy spillovers, rather than financial spillovers, were the largest single source of the increase in international macro-financial spillovers observed during the start of the COVID-19 pandemic.

The final dimension of our empirical analysis is motivated by past work In the literature on systemic risk such as Allen et al. (2012),Giglio et al. (2016) and Brownlees and Engle (2017), who have found empirical evidence that systemic risk measures have forecasting ability for future macroeconomic conditions. Motivated by these studies, we perform a detailed empirical analysis to see if the same is true for our macro-financial spillover measures. A key advantage of our DY-based approach in this context is that it directly provides a set of pairwise directional spillover measures rather than just a single higher-level numerical measure. This allows us to straightforwardly construct combination forecasts from these individual pairwise measures, an approach which has been found to provide gains in predictive ability relative to forecasts based on individual predictors elsewhere in the finance literature, such as Rapach et al. (2010) and Paye (2012).

We find that our measures provide consistently stronger predictive ability for macroeconomic conditions across a range of forecast horizons than existing systemic risk measures, including the CATFIN and SRISK measures proposed by Allen et al. (2012) and Brownlees and Engle (2017) respectively. This outperformance over existing systemic risk measures is especially pronounced in an out-of-sample context or when forecasting more specific aspects of macroeconomic conditions rather than aggregate macroeconomic conditions. Our measures' improvements in predictive accuracy are substantial during the 2008–2009 crisis.

The remainder of the paper is organised as follows. Section 2 introduces the methodology we develop for quantitatively measuring the strength and structure of macro-financial spillovers. Section 3 applies our methodology to perform a detailed empirical analysis of macro-financial spillovers in the US and a more concise analysis for a set of five other advanced economies. Section 4 examines the predictive ability of our macro-financial spillover measures for future macroeconomic conditions and finally Section 5 concludes.

2. Macro-financial spillover estimation

Our quantitative measures of spillovers are based on the established DY spillover methodology first proposed by Diebold and Yılmaz (2009) and developed further by Diebold and Yılmaz (2012, 2014)[2], which has been applied extensively to study financial spillovers but has not been employed in a macro-financial context. The DY approach relies on the forecast error variance decomposition from a standard vector autoregressive (VAR) model, in which all series are observed at a common sampling frequency. However, the combinations of the real economy and financial time series of interest here will typically contain series at different sampling frequencies, with financial series generally available at much higher sampling frequencies than macroeconomic series.

The traditional solution would be to aggregate all high-frequency financial series to the sampling frequency of the lowest frequency real economy series employed before applying the standard DY methodology to the transformed data. Whilst simple, the obvious drawback of such an approach is the potentially relevant information lost when aggregating the higher frequency financial series. Instead, we avoid these issues by replacing the standard common-frequency VAR model with a mixed-frequency VAR (MF-VAR) model. This allows us to employ high and low-frequency series together, estimating our spillover measures directly from a mixed-frequency macroeconomic and financial dataset.

It is worth noting that, as with the works on financial spillovers and systemic risk measurement pointed out in the Introduction, our DY-based approach to spillover estimation does not make any assumptions about the source of these macro-financial spillover channels or their theoretical underpinnings. Whilst more theoretical or structural approaches have potential advantages in other contexts, they are problematic for our objective of analysing macro-financial spillovers at an economy-wide level given the difficulties in formulating a structural model that is simultaneously plausible and tractable for this goal.

2.1. The mixed-frequency VAR model

We follow Ghysels (2016) for the specification and estimation of the MF-VAR. To illustrate the approach, we assume for simplicity that there are only two distinct sampling frequencies (high-frequency and low-frequency). We also assume that the number of high-frequency time periods is the same in each low-frequency period. Both of these are true for the empirical analysis here. However, it should be noted that the methodology is generally applicable and that these assumptions can be relaxed at the cost of more complex notation and implementation.

Formally, we observe a K-dimensional mixed-frequency vector process, which contains $K_L < K$ low-frequency macroeconomic or real economy series and $K_H = K - K_L$ high-frequency financial series. In terms of low-frequency time periods, which we index by τ_L, the low-frequency macroeconomic series are observed once per period and collected in the K_L-dimensional vector process $x_L(\tau_L)$. Each high-frequency financial series is observed m times every low-frequency time period. We group the high-frequency observations within each low-frequency period by series, with the m-dimensional vector $x_{H,i}(\tau_L)$ containing the m values for the i-th high-frequency series that are observed in low-frequency time period τ_L. In total we therefore have K_H m-dimensional vectors $x_{H,1}(\tau_L), \ldots, x_{H,K_H}(\tau_L)$ of high-frequency observations in each low-frequency time period.

We create a stacked vector for each low-frequency time period that contains both the K_L-dimensional vector of low-frequency real economy observations, $x_L(\tau_L)$, and the K_H m-dimensional vectors containing all the high-frequency financial data observed during the same low-frequency time period. The resulting stacked vector is denoted by $\underline{x}(\tau_L)$ and is of dimension K_x, where $K_x \equiv (mK_H + K_L)$:

$$\underline{x}(\tau_L) \equiv \left[x_{H,1}(\tau_L)', \ldots, x_{H,K_H}(\tau_L)', x_L(\tau_L)' \right]'$$

Following Ghysels (2016), we then specify a standard VAR model for the stacked mixed-frequency vector $\underline{x}(\tau_L)$.The general form of the p-th order MF-VAR is thus given by:

$$\underline{x}(\tau_L) = A_0 + \sum_{j=1}^{p} A_j \underline{x}(\tau_L - j) + \underline{\varepsilon}(\tau_L) \tag{1}$$

where A_0 is an K_x-dimensional parameter vector, $A_j, j = 1, \ldots p$ are $(K_x \times K_x)$ parameter arrays and $\underline{\varepsilon}(\tau_L)$ is an K_x-dimensional vector of errors. Despite the somewhat non-standard composition of the vector $\underline{x}(\tau_L)$, the model is mathematically equivalent to a standard VAR. As such, standard methods for estimation and analysis of VAR models can be employed.

In addition to the stacked mixed-frequency vector process $\underline{x}(\tau_L)$ introduced above, we will also consider the associated K-dimensional low-frequency vector process denoted by $\overline{x}(\tau_L)$, which contains both the K_L low-frequency real economy series and the K_H high-frequency financial series appropriately aggregated down to the lower frequency of the real economy series:

$$\overline{x}(\tau_L) \equiv \left[x_{HtL}(\tau_L)', x_L(\tau_L)' \right]'$$

[2] Numerous extensions to the original DY spillover methodology have also been proposed, with notable examples including factor-augmented VARs (Claeys and Vašíček, 2014), asymmetries in spillovers (Barunik et al., 2015), frequency-domain methods (Barunik and Křehlík, 2018), and quantile connectedness (Ando et al., 2022).

where $x_{HtL}(\tau_L)$ is used to denote the set of high-frequency series aggregated to the lower frequency in the τ_L-th time period. We primarily employ return levels as financial series in the current work, with return volatilities also considered in the Appendix. As such, the high-frequency financial series contained in the vectors $x_{H,1}(\tau_L), \ldots, x_{H,K_H}(\tau_L)$ will be weekly returns (or return volatilities) and those in the aggregated low-frequency vector $x_{HtL}(\tau_L)$ are monthly returns (or return volatilities). We also specify a standard common-frequency VAR model for the low-frequency vector process $\bar{x}(\tau_L)$, which we will refer to as the common-frequency VAR (CF-VAR).

While the MF-VAR and CF-VAR are both technically specified at the lower sampling frequency, the MF-VAR also incorporates higher frequency information available within each low-frequency time period that is not used by the CF-VAR. We use a combination of monthly macroeconomic series and weekly financial series, allowing the MF-VAR to incorporate potentially relevant intra-month information on financial market behaviour at the weekly frequency. In the case of the CF-VAR, this high-frequency information is discarded when the financial series are aggregated down to the monthly frequency. As noted during the introduction, Ghysels (2016),Schorfheide and Song (2015) and others have previously shown that the use of mixed-frequency methods may provide gains in accuracy for both estimation and forecasting in the context of VAR models relative to a common-frequency approach. During our empirical analysis, we will directly compare the DY spillover measures obtained from the traditional CF-VAR with those from the MF-VAR to investigate the impact of including this additional high-frequency information.

2.2. Forecast error variance decomposition and spillover measures

After the specified VAR model has been estimated, the next step when computing the DY spillover measures is to compute the forecast error variance decomposition (FEVD) arrays for the VAR. Following Diebold and Yılmaz (2012), we employ the approach of Pesaran and Shin (1998) to compute generalised FEVD values. This approach is widely employed in the literature, and so numerical details are relegated to Appendix A.1.

For a generic K-dimensional VAR, the FEVD arrays are of dimension $(K \times K)$ and of the form:

$$\begin{bmatrix} \phi_{11}(H) & \cdots & \phi_{1K}(H) \\ \vdots & \ddots & \vdots \\ \phi_{K1}(H) & \cdots & \phi_{KK}(H) \end{bmatrix} \quad \text{for} \quad H = 1, 2, \ldots \tag{2}$$

where $\phi_{kl}(H)$ for $k, l = 1, \ldots, K$ is the fraction of the H-step-ahead error variance in forecasting series k that is attributable to shocks in series l. The FEVD array elements thus have a clear interpretation as measures of spillovers and shock transmission between the series in the system. More specifically, the pairwise DY spillover from series i to series j is given by:

$$S_{ij}(H) = \frac{100}{K} \cdot \phi_{ji}(H) \tag{3}$$

Multiplying the relevant FEVD element $\phi_{ji}(H)$ by the factor $100/K$ ensures that each pairwise spillover value is expressed as a percentage of the total forecast error variance across all series in the VAR[3].

The DY spillover measures are complementary to alternative systemic risk measures as tools for monitoring market conditions. It is worth emphasising again that the DY pairwise spillover measures are directional in the sense that $S_{ij} \neq S_{ji}$ for $i \neq j$. This is a key theoretical advantage compared to most common measures of pairwise association such as correlation, which are non-directional and measure only the strength of association between two series. Indeed this is also an important difference between the DY spillover measures and other established measures of systemic risk employed in the finance literature, such as the CATFIN measure of Allen et al. (2012) or the SRISK measure of Brownlees and Engle (2017). While some of these measures have a directional aspect, the nature or scope of the directionality measured is restricted by construction. For example, SRISK, when computed at the firm or market level, corresponds to the contribution of that firm or market to the overall systemic risk within the system conditional on a systemic market decline occurring.

The fact that the DY approach naturally produces a set of pairwise spillover measures rather than simply a single numerical measure also emerges as an important advantage when we employ the measures to forecast macroeconomic conditions in Section 4. As demonstrated later, the fact that each of the directional pairwise spillover measures contains different information concerning the structure of macro-financial spillovers allows us to exploit the use of combination forecasts, which have had much success in the previous literature on predictability.

Whilst the pairwise spillover measures of Eq. (3) permit a detailed analysis of the direction and structure of spillovers, more aggregated measures may also be helpful to quantify the overall strength of spillovers concisely. We, therefore, employ both the disaggregated pairwise measures above and the most aggregated measure proposed by Diebold and Yılmaz (2012). This is referred to as the total spillover index and provides a single numerical measure of the overall level of spillovers between the series included in the underlying VAR. The measure is thus similar in spirit to many of the economy-wide systemic risk measures proposed elsewhere in the literature, such as CATFIN and the aggregate variants of SRISK and CoVaR. The total spillover index is computed as:

[3] This follows because, as discussed in Appendix A.1, the sum of the elements in each row of the FEVD array equals one, giving a sum over all elements equal to K.

$$S(H) = \frac{100}{N} \sum_{\substack{i,j=1 \\ i \neq j}}^{K} \phi_{ji}(H) \tag{4}$$

The total spillover index gives the percentage of the total H-step-ahead forecast error variance for all series that is attributable to shocks *across* series, i.e. excluding the direct effect of shocks to each given series on itself.

As shown by Diebold and Yılmaz (2014), parallels can be drawn between the DY spillover measures and other systemic risk measures proposed in the literature, such as the CoVaR measure of Adrian and Brunnermeier, 2016 and the MES measure that features in the work of both Acharya et al. (2016) and Brownlees and Engle (2017). Besides the slightly different aspect of connectedness being measured by the different approaches, the key difference is that CoVaR, MES and the majority of the other systemic risk measures in the literature focus on tail risk or connectedness conditional on the market or financial institution being in an adverse state. The DY approach, on the other hand, measures unconditional connectedness in the sense that it considers the average or mean state of the market. This average state is then allowed to change over time through the use of a dynamic estimation environment as discussed in Section 3.

2.3. Transformation of mixed-frequency forecast error variance decomposition

The FEVD arrays for the MF-VAR model are computed as in the common-frequency case. However, they will have a non-standard structure arising from the non-standard composition of the stacked vector, $\underline{x}(\tau_L)$. Whilst the standard DY spillover measures can be computed directly from these mixed-frequency FEVD arrays, the interpretation of the measures obtained will differ from the standard common-frequency case.

The relevant issues and concepts are best illustrated using a simple example that is closely related to the empirical analysis in the following sections. Specifically, we employ a bivariate model with one low-frequency monthly macroeconomic series and one high-frequency financial weekly series for exposition. We thus have $m = 4, K_L = 1 \text{and} K_H = 1$, giving a stacked mixed-frequency vector of dimensions $K_x = 5$, with the form $\underline{x}(\tau_L) = [x_H(\tau_L, 1), \ldots, x_H(\tau_L, 4), x_L(\tau_L)]'$. For the corresponding common-frequency VAR, we have a (2×1) vector process $\overline{x}(\tau_L) = [x_{HtL}(\tau_L), x_L(\tau_L)]'$. This results in (5×5) FEVD arrays for the MF-VAR and (2×2) arrays for the CF-VAR, given respectively by:

$$\begin{bmatrix} \theta_{11}(H) & \ldots & \theta_{15}(H) \\ \vdots & \ddots & \vdots \\ \theta_{51}(H) & \ldots & \theta_{55}(H) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \phi_{11}(H) & \phi_{12}(H) \\ \phi_{21}(H) & \phi_{22}(H) \end{bmatrix} \quad \text{for} \quad H = 1, 2, \ldots \tag{5}$$

with the differences in notation used only to distinguish the FEVD elements for the MF-VAR and CF-VAR.

It is clear that the FEVD arrays for the MF-VAR in (5) will be larger than those for the corresponding CF-VAR, since $K_x > K$. This arises because the weekly high-frequency series observed in each low-frequency monthly time period are treated mathematically as separate series when estimating the MF-VAR but enter the CF-VAR as a single monthly series. As a result, in the common-frequency case a single FEVD element completely characterises the directional spillovers at the chosen forecast horizon between a given pair of macroeconomic or financial series. In contrast, in the mixed-frequency case, it will generally be characterised by multiple FEVD array elements.

We, therefore, develop an approach for transforming the FEVD arrays obtained from the MF-VAR to produce new FEVD arrays with the same structure and dimensions as those for the corresponding CF-VAR. We can then compute DY spillover measures from these transformed arrays that are directly comparable to those in the standard common-frequency case. The basic intuition of the transformation approach is outlined here, with mathematical details found in Appendix A.2.

Intuitively we exploit the correspondence between the elements of the FEVD arrays for the mixed-frequency and common-frequency cases. More specifically, for the current example we group the MF-VAR FEVD elements in (5) into sub-arrays as follows:

$$\begin{bmatrix} \Theta_{11}(H) & \Theta_{12}(H) \\ \Theta_{21}(H) & \Theta_{22}(H) \end{bmatrix} \quad \text{for} \quad H = 1, 2, \ldots \tag{6}$$

where:

$$\Theta_{11}(H) \equiv \begin{bmatrix} \theta_{11}(H) & \ldots & \theta_{14}(H) \\ \vdots & \ddots & \vdots \\ \theta_{41}(H) & \ldots & \theta_{44}(H) \end{bmatrix} \qquad \Theta_{12}(H) \equiv \begin{bmatrix} \theta_{15}(H) \\ \vdots \\ \theta_{45}(H) \end{bmatrix}$$

$$\Theta_{21}(H) \equiv [\, \theta_{51}(H) \quad \ldots \quad \theta_{54}(H) \,] \qquad \Theta_{22}(H) \equiv \theta_{55}(H)$$

Each of the sub-arrays $\Theta_{kl}(H)$ in (6) can be viewed as a mixed-frequency analogue of the corresponding scalar element $\phi_{kl}(H)$ from the CF-VAR FEVD array in (5). For example, the (4×1) sub-vector $\Theta_{12}(H)$ characterise the effects of shocks to the monthly low-frequency series (series 2) on the weekly high-frequency series (series 1). Specifically, θ_{i5} for $i = 1, \ldots, 4$ mea-

sures the fraction of the H-step-ahead error variance in forecasting the high-frequency series in week i of the month that is attributable to shocks in the low-frequency series. The scalar element $\phi_{12}(H)$ for the common-frequency case describes the same directional pairwise relationship for the case where both series are observed at the lower monthly frequency.

The approach detailed in Appendix A.2 transforms each of the sub-arrays $\Theta_{kl}(H)$ in (6) into a scalar value, whose interpretation is directly comparable with the corresponding element $\phi_{kl}(H)$ in the standard common-frequency case. This comparability of the values is ensured by directly basing the transformation used on the mathematical definition of the generalised FEVD elements.

3. Macro-financial spillovers

Our empirical analysis of macro-financial spillovers focuses primarily on the United States, given the central role that the country plays in global economic and financial markets. We examine over four decades of data ending in June 2020, thus covering many significant economic and financial events, including the 2008–2009 global financial crisis and the large shocks resulting from the onset of the COVID-19 pandemic. In Section 3.3 we also more briefly examine the international dimension of macro-financial spillovers using data for five other advanced economies, namely Canada, France, Germany, Japan and the United Kingdom.

Section 3.1 begins by describing the data and implementation of the approach employed to estimate spillovers. Section 3.2 presents a graphical analysis of the US macro-financial spillover measures obtained using our approach. In this subsection, we also include measures obtained from the existing common-frequency approach to assess the practical effects of ignoring the high-frequency financial data when using a more traditional modelling approach.

3.1. Data and spillover estimation

On the real side of the economy, our series of interest is the Chicago Fed National Activity Index (CFNAI) of the Federal Reserve Bank of Chicago.[4] The CFNAI is employed at its regular monthly frequency and is constructed to have a mean value of zero, with positive (negative) values corresponding to growth above (below) its historical trend. We work with the level of the CFNAI series, given that this corresponds in a broad sense to the change in macroeconomic conditions[5].

On the financial side, we focus on equity and bond markets, represented by the S&P500 equity index and the 10-year US Treasury Note, respectively. Although the strength and structure of spillovers between bond markets and the real economy may vary with bond maturity and type (as suggested by Brenner et al., 2009), for simplicity, we restrict our attention to US sovereign bonds, specifically the 10-year Treasury Note.

We focus primarily on return levels for the financial series, assuming that changes in macroeconomic conditions will be linked to changes in asset values. However, we do repeat the core analysis of Section 3.2 using return volatilities for the sake of completeness, with the results found in Appendix C.2. Whilst the switch to return volatilities does induce some changes in the levels and dynamics of the various spillover indexes, the key empirical findings for return levels follow through primarily unchanged.

For the non-US analysis of Section 3.3, we use month-on-month industrial production growth as our macroeconomic variable for the real side of the economy given the lack of an equivalent index to the CFNAI for other countries. To provide results that are entirely comparable across countries, we also repeat the US analysis using monthly IP growth in place of the CFNAI. Similar to the US, for the financial side of the economy, we use return levels for the major national equity index for each country[6] and generic national 10-year government bond indexes obtained from Bloomberg.

For the US analysis in the current section, our sample period spans 1975:01 to 2020:06, thus including many significant economic and financial events of recent decades. Due to limited data availability for the non-US government bond index series, the sample period for the international analysis begins in 1990:01 and ends in 2020:04.

Whilst the inclusion of the massive shock associated with the COVID-19 pandemic is of much interest; it has been noted by recent work such as Lenza and Primiceri (2020) and Schorfheide and Song (2020) that its sheer magnitude leads to challenges for parameter estimation and forecasting using standard econometric methods. This is relevant both for the estimation of the underlying VAR model used to compute the spillover measures and also the predictive regressions used for the forecasting exercise in Section 4. Given the size of the shock and its unique nature, these studies have suggested that it should be explicitly considered at the estimation stage, such as using some form of down-weighting. We do not pursue such approaches given that it is not the focus of the current work, but we acknowledge the challenges posed for standard econometric methods such as those we employ. Notwithstanding these issues, analysing spillovers between the real economy and financial markets during this period is of interest. Furthermore, given the rolling window approach used to estimate the spil-

[4] CFNAI is frequently used in empirical work to provide a single numerical measure of US macroeconomic activity that is broader and less noisy than specific series such as industrial production (see for example Allen et al., 2012). It is a composite index derived from 85 underlying macroeconomic indicators grouped into four categories: production and income, employment unemployment and hours, consumption and housing, and sales, orders and inventories.

[5] We also repeated the analysis performed in the current section using the four CFNAI subcomponents in place of the aggregate CFNAI index. These results are not reported here to conserve space but are available upon request. These CFNAI subcomponents are, however, employed during the forecasting exercise of Section 4, with further details of their characteristics found in the associated discussion.

[6] S&P/TSX for Canada, CAC40 for France, DAX for Germany, Nikkei 225 for Japan, FTSE100 for the UK and as before the S&P500 for US.

lover indexes, this issue affects only the spillover values obtained for the pandemic time periods. Therefore, in the current section, we choose to include these observations in our analysis.

The raw data for the two financial series consists of daily closing prices, from which we produce a closing price series at a weekly frequency. To sidestep the practical issues caused by the variation in the number of weeks per month, we employ a data pre-processing and transformation approach to the daily series to produce weekly series with a constant four weeks per month. These weekly closing prices are used to compute weekly returns. Further details of the data processing approaches, together with plots of all the US series employed, can be found in Appendix B. It is interesting to note particular events associated with substantial movements and volatility in the series, including the 1980–1981 recession, the Asian and Russian financial crisis, 9/11, the 2008–09 global financial crisis and the recent onset of COVID-19. We will discuss the associated spillovers from these and other events in more detail below.

Our interest lies in obtaining dynamic estimates of macro-financial spillovers, rather than static full-sample estimates, to investigate how the strength and structure of spillovers have varied over time. To achieve this, we use a standard rolling window estimation approach in which the parameters of the MF-VAR and the connectedness measures are re-estimated for each window. A window length of 60 months is employed since it appears to offer a good balance between providing a sufficient sample size to estimate the parameters of the underlying MF-VAR to an appropriate level of accuracy and allowing dynamics of connectedness to be captured. We have, however, checked our results' robustness to reasonable changes in the window length, and there is no qualitative impact on the results[7]. When computing the spillover measures, we primarily considered forecast horizons of 3, 6 and 12 months, consistent with most previous studies. However, we found that the estimates did not show significant sensitivity to the choice of the forecast horizon, and so reported results only for the horizon of 3 months.

3.2. US macro-financial spillovers

We begin in Fig. 1 by plotting the total spillover indexes obtained from both the mixed-frequency and common-frequency approaches. It can be seen that the estimated total spillover indexes obtained from the two methods show broadly the same movements over the sample period. However, despite the high correlation between the indexes, the level of total macro-financial spillovers implied by the new mixed-frequency approach is, with one or two exceptions, consistently higher than that obtained from the common-frequency approach. For example, the average values of total spillovers for the mixed-frequency and common-frequency approaches are 25.79% and 15.12%, respectively, representing the proportion of the total forecast error variance in the entire system due to shocks across series. Thus by aggregating the financial data to monthly frequency and ignoring the additional intra-monthly information it contains, one obtains substantially lower estimates of the level of spillovers across the real and financial sectors.

A possible explanation for this finding of higher average spillover levels for the mixed-frequency case can be found in the previous literature on the effects of macroeconomic announcements on financial markets. Studies such as Andersen et al. (2003) and Green (2004) have employed high-frequency intraday financial data to study the effects of macroeconomic news and announcements over short time periods and have found significant intraday effects. However, they note that lower frequency daily data prevents these effects from being observed and thus may bias estimates of the response in the financial markets downwards. Intuitively an analogous explanation can be applied in the current analysis, in which the arrival of intra-month shocks to either the financial or real series may result in significant within-month spillovers that are visible through the use of weekly data for some series in the mixed-frequency case, but are either ignored entirely or underestimated when using purely monthly data in the common-frequency approach.

Considering the dynamics of the total spillover indexes briefly, we see substantial fluctuations over the sample period, many of which coincide with major economic or financial events. Some notable examples are marked in Fig. 1, with the most significant spikes in spillovers occurring during the 2008–2009 global financial crisis, particularly around the collapse of Lehman Brothers, the bailout of AIG and Fannie Mae and Freddie Mac being placed in government conservatorship.

As can be seen in Fig. 1, the most significant spike in macro-financial spillovers during our sample period has occurred with the onset of the COVID-19 pandemic, dominating even the 2008 global financial crisis. Between February and March 2020, the common-frequency and mixed-frequency estimates rose from 16 to 37 and from 25 to 44, respectively. This is expected given that the shock resulting from the COVID-19 pandemic has been the largest negative shock to hit the global economy since the Great Depression of 1929.

In addition, except the global financial crisis of the last quarter of 2008, April-May 2010, and the COVID-19 pandemic, the movements in the mixed-frequency spillover measure are typically substantially smoother than the common-frequency spillovers, which frequently exhibits significant upward or downward jumps such as those in October 1987, September 1992 and August 1998. This likely results from incorporating intra-month financial information in the mixed-frequency analysis, which results in the effects of sustained shocks to financial series being spread across consecutive weeks and gradually incorporated into the spillover index. On the other hand, in the common-frequency case, only the accumulated shock is observed at the end of the month, leading to a more significant jump in the index when this information is incorporated into the new value of the index.

[7] This point is illustrated in Appendix C.3 with equivalent plots to Fig. 2 using window sizes of 45 and 90 months.

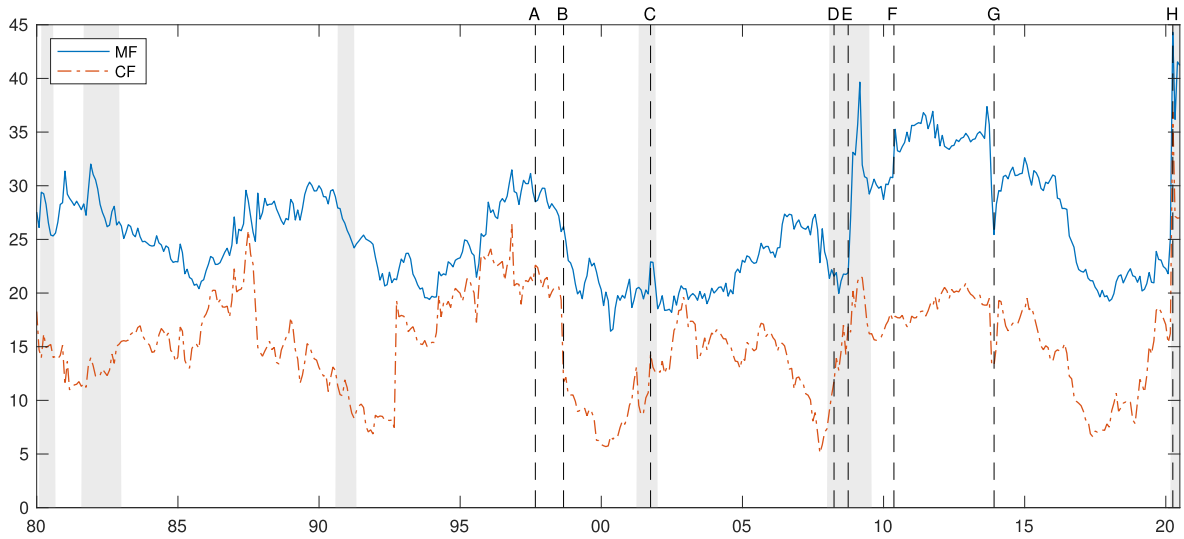

Fig. 1. Total spillover indexes between the financial and real economy series. Total spillover indexes for mixed-frequency (denoted MF) and common-frequency (denoted CF) approaches are presented for the sample period 1980:01 to 2020:06. Return levels are employed for the financial S&P500 and 10-year Treasury Note series, and levels for the real economy CFNAI series. Values are computed using a 3-month forecast horizon and a 60-month rolling window. The points marked are as follows. A: Asian financial crisis, Jul '97, B: Russian financial crisis and LTCM collapse, Aug to Sept '98, C: September 11, Sept '01, D: the collapse of Bear Stearns, Mar '08, E: Lehman Brothers collapse, AIG bailout and Fannie Mae and Freddie Mac being placed in government conservatorship, May '09, F: start of the EU debt crisis in April '10, flash crash of May '10, G: tapering of QE in December 2013, and H: COVID-19 pandemic, March '20. Shaded areas correspond to NBER US recession dates..

The total spillover index provides an informative but highly aggregated measure that potentially hides many exciting details of the structure of macro-financial spillovers. We decompose the total spillover measures of Fig. 1 into their component pairwise directional spillover measures. We first represent this information in the form of spillover decomposition plots in Fig. 2. Given that the total spillover index equals the sum of all pairwise measures, the top of the complete shaded area corresponds to the relevant total spillover index (as previously plotted in Fig. 1), and the shaded areas beneath represent the contribution of each pairwise spillover. These spillover decomposition plots thus provide an effective way to visualise the relative importance of each directional spillover channel in total spillovers over time. Panels (a) and (b) present the mixed-frequency and common-frequency spillover decomposition plots for the entire sample period, whereas panels (c) and (d) span only 2008:06 to 2020:12, allowing more precise analysis of the 2008–2009 crisis and the start of the COVID-19 pandemic.

For the mixed-frequency case in panels (a) and (c), it is immediately apparent that the vast majority of the (total) spillovers originates in financial markets, represented by the bottom four shaded areas of the spillover decomposition plots. The contribution of the real side of the economy (the sum of the areas represented by the yellow and orange colours) accounts for only a tiny part of the total spillovers in this context, whereas for the common-frequency case in panels (b) and (d) they are larger. Whilst there is a clear difference in the average levels of real to financial spillovers over the sample period for mixed-frequency and common-frequency (1.94% and 2.69% respectively), this difference is most notable during the late 1980s, during the 2008–2009 financial crisis and the COVID1-9 pandemic (see panels (c) and (d)), and to a lesser extent in the early 1980s and late 1990s.

As discussed above, we hypothesise that this difference is due to the common-frequency index not incorporating higher frequency financial information. Thus the estimates obtained suggest a relatively more minor role for financial markets. This is expected given that financial markets can respond much more quickly to either economic or financial shocks than real economy variables. The incorporation of weekly financial data allows our mixed-frequency approach to detect these high-frequency intra-month responses in financial markets, which would be ignored by the common-frequency approach using only monthly information.

The pairwise spillover measures can also be presented as standard line plots. This complements the spillover decomposition plots of Fig. 2, with the latter being more suited to analysing the relative contribution of the individual spillover measures to the total level, and the former more suited to exploring the dynamics of specific spillover measures in absolute terms.[8]

[8] To conserve space, the line plots and the associated discussion are included in Appendix C.1, where we examine how changes in pairwise spillovers relate to significant historical events as we did previously for the total spillover index in Fig. 1.

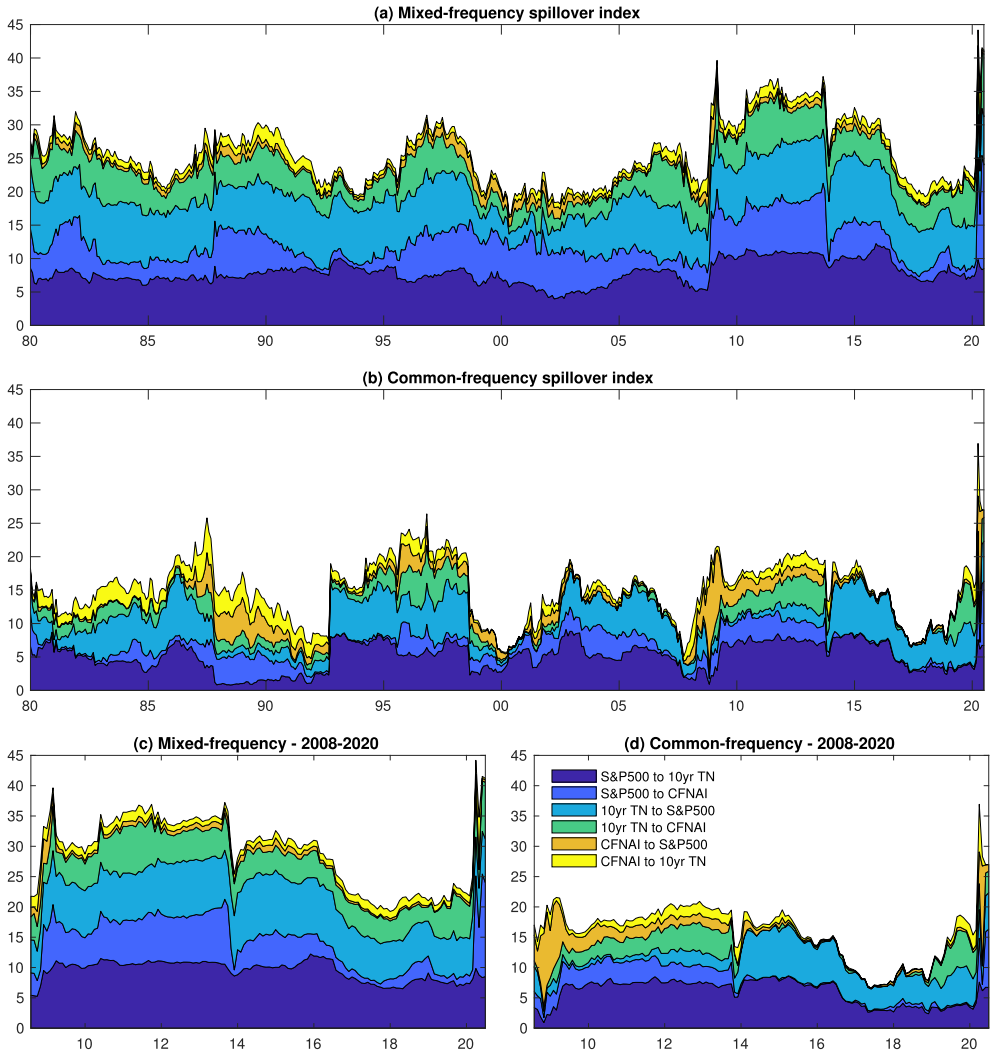

Fig. 2. Decomposition of total macro-financial spillover indexes into pairwise components. The figure presents area plots in which the top of the complete shaded area corresponds to the relevant total spillover index and each shaded area below representing the contribution of each specific pairwise spillover to the value of the total spillover index. The relevant spillover measures for the mixed-frequency case are plotted in panels (a) and (c) and those for the common-frequency case in panels (b) and (d). Return levels are used for the financial S&P500 and 10-year Treasury Note series and levels for the real economy CFNAI series. Values are computed using a 3-month forecast horizon and a 60-month rolling window for the sample period 1980:01 to 2020:06 in panels (a) and (b) and 2008:06–2020:06 in panels (c) and (d).

3.3. International macro-financial spillovers

The international dimension of macro-financial spillovers is clearly of interest given the connectedness of modern financial and economic markets across national borders. This is an extensive topic for possible future research, but here we take an initial step towards a fuller examination of international macro-financial spillovers.

We perform a series of two-country analyses consisting of the US and one other advanced economy, considering in turn Canada, France, Germany, Japan and the UK. For each estimated two-country VAR model, we include on the financial side weekly returns for the S&P500 and the respective major non-US national equity index detailed previously in footNote 5. On the real side, we employ monthly IP growth for the US and non-US advanced economy due to the data limitations previously noted in Section 3.1. These four variables allow us to examine directional cross-country financial-to-financial, real-to-real, financial-to-real and real-to-financial spillovers, without the number of pairwise spillover measures becoming impractically large for discussion.

The practical issue encountered is that common global shocks drive some of the movements in variables of the same type across countries e.g. the S&P500 and FTSE100. This makes it difficult in some periods to separate out the spillover effects of, for example, shocks to the S&P500 on UK IP growth and shocks to the FTSE100 on UK IP growth. To minimise this, we restrict

the error covariance matrix of the estimated VAR such that shocks that are both cross-country *and* cross-variable, for example US IP growth to UK equity returns, are contemporaneously uncorrelated. Correlation between shocks within each country (e.g. UK IP growth and UK equity returns) and shocks to the same variable type for different countries (e.g. US and UK IP growth) are unrestricted.

In terms of spillovers, this implies that *contemporaneous* (i.e. 0-step-ahead) pairwise spillovers for these cross-country cross-variable combinations are zero, leaving any cross-country contemporaneous spillovers to occur via the financial-to-financial or real-to-real channels. Whilst the restrictions imposed do impact the spillover measures obtained at longer horizons of $H \geqslant 1$, which are the object of interest here and in most other DY-based analyses, these spillovers will be non-zero unlike those for $H = 0$. To implement these restrictions a feasible GLS estimation approach is used. All other aspects of the methodology are identical to that discussed previously.

This two-country four-variable setup allows us to analyse key cross-country macro-financial spillovers without the set of spillover measures to be discussed becoming impractically large for the current subsection. It also has the benefit of minimising the parameter proliferation issues encountered in larger VAR models. As demonstrated by Demirer et al. (2018), this issue can be solved in the context of DY spillover measures even for extremely large VARs by using shrinkage based methods such as LASSO, but this would take us beyond the scope of the current work.

We present the mixed-frequency and common-frequency total spillover indexes for the five pairs of countries in Fig. 3. Similar to the single country US case, whilst the total spillover indexes based on the mixed-frequency and common-frequency approaches tend to follow similar trajectories over time in each country, the values for the mixed-frequency approach are almost always higher than the corresponding index based on the common-frequency approach. For conciseness, the remainder of the discussion below refers to the mixed-frequency indexes only.

During the 2008–09 global financial crisis the total spillover indexes for all country pairs increased significantly to values within the range 40–45%. For the US-Germany pair, total spillovers continued to rise further late 2014, presumably due to a combination of the effects of the EU sovereign debt crisis and Germany's strong economic and financial links with US and world markets. Interestingly however, whilst total international spillovers for the US-France and US-UK pairs also remain high during the same period, they do not show the same sustained increase as the case of Germany. For the UK this can be attributed to its lack of membership of the Eurozone, but the lack of a similar rise in France is more difficult to explain outside of lower economic and financial integration with US markets. Total spillovers for the US-Japan pair begin to drop much quicker from 2011 onwards, due to the relative isolation from the turbulence occurring in the EU.

For all country pairs, the response of the mixed-frequency index to the shock caused by the beginning of the COVID-19 pandemic in Spring 2020 was broadly similar to that observed in the US, with a very sharp elevation of total spillovers. We observe that for US-France and US-Germany, the peak in the total spillover index occurred in March 2020, whereas in the remaining country pairs, the current peak is recorded in April. This timing is consistent with the spread of the pandemic, given that most large mainland European countries such as France and Germany imposed strict lockdown measures earlier than the other countries, including the UK. The size of the increases relative to their February 2020 levels are similar across countries, typically in the range of 20–25 percentage points. However, there is some variation that appears related to country-level variation in the severity and timing of the pandemic, and the scale of measures imposed to limit its spread in the non-US country. In particular, the largest increase of approximately 27 percentage points is observed for the US-France pair, and the lowest of 19 percentage points for the US-Japan pair, which is consistent with the level of cases and restrictions imposed in these countries at the time. As noted in the US analysis, extending the sample period would allow international macro-financial spillovers during later stages of the pandemic to be studied, but this is not the primary focus of the current analysis and so is left for future work.

Next, we examine international macro-financial spillover structure in more detail using a subset of the pairwise spillover measures from which the total index is obtained[9]. To conserve space we restrict attention to three pairs of countries, namely US-Germany, US-Japan and US-UK, with results for Canada and France being similar and available upon request. Fig. 4 plots the relevant cross-country or international pairwise spillovers for the selected country pairs.

Beginning with the relative importance of the various pairwise spillover types, as was true in the previous US analysis, the majority of total macro-financial spillovers are comprised of spillovers originating in financial markets, represented by the top two rows of subplots in Fig. 4. International equity to equity market spillovers in the second row are substantially higher than other spillover sources on average across the sample period, typically within the range of 6% to 10%.

Although lower on average, cross-country spillovers from equities to the real sector in the first row are also large relative to other pairwise spillover types during the 2008–09 global financial crisis, frequently reaching sustained levels of 4–6%. International financial to real spillovers also clearly pick up the early stages of the COVID-19 pandemic at the end of the current sample period, with large spikes observed for both the US-Germany and US-UK pairs and a smaller peak for US-Japan, consistent with the lower case numbers and level of restrictions in Japan at the time.

In this first row, spillovers from non-US equities to US IP exceed those from US equities to non-US IP in all three cases in the period preceding the 2008–09 global financial crisis. Given the global importance and size of US financial markets, it might be assumed that spillovers from US financial to non-US real variables should naturally be larger than those from

[9] We do not produce spillover decomposition plots analogous to Fig. 2 because with four series in the VAR we obtain 12 pairwise measures rather than the previous 6, making the decomposition plot difficult to interpret.

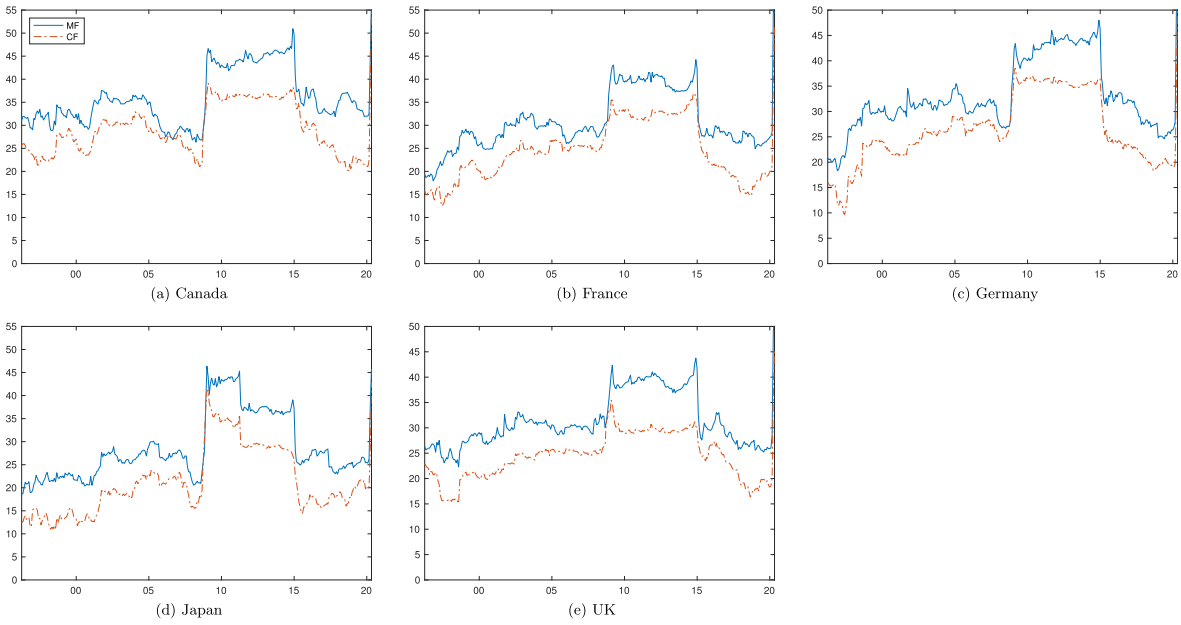

Fig. 3. International total spillover indexes between the financial and real economy series. The figure presents total spillover indexes for mixed-frequency (denoted MF) and common-frequency (denoted CF) approaches for the sample period 1995:01 to 2020:04. Return levels are employed for the financial series, using the S&P500 and a major equity index for the chosen non-US country (S&P/TSX for Canada, CAC40 for France, DAX for Germany, Nikkei 225 for Japan and FTSE100 for the UK). For the real economy series, month-on-month growth rates of industrial production for the US and chosen non-US country are employed. Spillover index values are computed using a 3-month forecast horizon and a 75-month rolling window..

non-US financial markets to US IP. At least in the case of Germany and the UK, it may be the case that the DAX and FTSE100 equity returns are acting as a proxy for financial conditions within Europe more generally and thus viewing these as being Germany- or UK-specific spillovers is not entirely valid.

It should also be remembered that directional spillover magnitude depends not only on the importance of shocks from the transmitting variable, but also the responsiveness (or lack of responsiveness) of the receiving variable to those shocks. It may be that US industrial production in the period preceding the crisis was more sensitive to shocks in international financial conditions than IP in other countries. This could be due to the composition (both in terms of sectors and the importance of exports), or size of manufacturing output, which varies significantly across these four countries; in 2015 manufacturing output as a proportion of global manufacturing in Germany, Japan, the UK and the US was 7%, 10%, 2% and 18% respectively, making total US manufacturing output significantly higher than that of the other countries. Likewise, the small size of UK industrial production both as a proportion of national output and globally, provides a possible explanation for its lack of responsiveness to shocks in US financial conditions, particularly during the 2008–09 global financial crisis. It should be noted that while spillovers from the FTSE100 to US IP look high, they are actually around 4% during this episode, which is only slightly greater than those observed for Germany from the DAX to US IP.

Whilst cross-country real to equity spillovers are typically low and in the range of 0–1%, increases are nonetheless observed consistent with key events. The most consistently visible example is the 2008–09 crisis and subsequent recession, but the 2011 Tohoku earthquake and associated recession is also very clearly visible as a large increase in spillovers from Japanese IP to the S&P500. The start of the pandemic is again visible as an increase in real to financial spillovers, but to a lesser extent than the financial to real spillovers discussed above.

Finally, the international real to real spillovers in the final row of Fig. 4 have no direct analogue in the previous single country US analysis. For much of the sample period, cross-country spillovers between IP growth fluctuate within the 0–2% range. Increases in spillovers from US IP to German IP and to UK IP are visible during the global financial crisis, as is the impact of the 2011 Tohoku earthquake and recession on spillovers from Japanese IP to US IP. Perhaps most notably, during the start of the COVID-19 pandemic, real to real international spillovers showed larger spikes than any other spillover type despite their relatively low levels over the rest of the sample period. For the US-Germany and US-UK pairs, cross-country IP growth spillovers increased from close to zero up to 8% and 6% for the case of US-Japan as industrial output dropped dramatically with lockdowns and the sharp contraction in production and trade.

4. Macro-financial spillovers as predictors of future macroeconomic conditions

A reoccurring question of interest in the literature on systemic risk is whether the quantitative measures developed have predictive ability for future macroeconomic conditions. In particular, it is often suggested (see, for example Allen et al., 2012

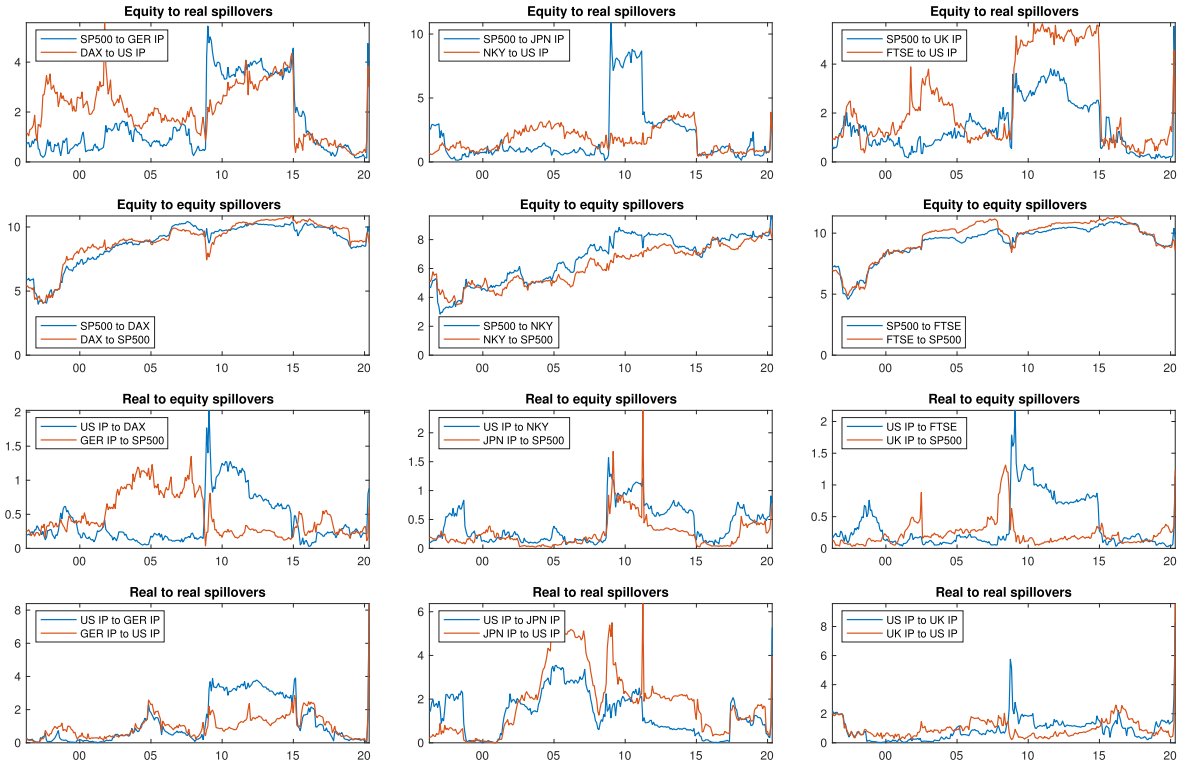

Fig. 4. International total spillover indexes between the financial and real economy series. The figure presents total spillover indexes for mixed-frequency pairwise spillover measures for the sample period 1995:01 to 2020:04. Return levels are employed for the financial series, using the S&P500 and a major equity index for the chosen non-US country (DAX for Germany, Nikkei 225 for Japan and FTSE100 for the UK). For the real economy series, month-on-month growth rates of industrial production for the US and chosen non-US country are employed. Spillover index values are computed using a 3-month forecast horizon and a 75-month rolling window..

and references therein) intuitively that an increase in systemic risk may hurt current and future economic conditions, primarily through a reduction in lending from banks to the non-financial sector. Allen et al. (2012),Giglio et al. (2016) and Brownlees and Engle (2017) amongst others have found empirically that various quantitative measures of systemic risk do have forecasting ability for essential macroeconomic series. Motivated by these findings, we perform a similar exercise to examine whether the current level of macro-financial spillovers also exhibits predictive power for future macroeconomic conditions.

4.1. Forecasting environment and predictive regressions

Our empirical approach closely follows those of Allen et al. (2012) and Brownlees and Engle (2017), and we focus on the problem of forecasting macroeconomic series, specifically the level of the CFNAI and its subcomponents, using predictive regressions of the form:

$$y_{t+n} = \alpha_n + \beta_n s_t + \sum_{i=0}^{q} \gamma_{n,i} y_{t-i} + \delta_n X_t + \epsilon_t \tag{7}$$

where y_t is the value of the macroeconomic series of interest, s_t is the value of the chosen macro-financial spillover measure and X_t is a vector of financial control variables commonly used in the literature as simple predictors of future macroeconomic conditions. The vector of control variables X_t consists of the current values of the default spread, the term spread, the return on the S&P 500 equity index and the credit-to-GDP gap[10]. The forecast horizon is denoted by n and we consider forecast horizons from one month ($n = 1$) up to a maximum of 12 months ($n = 12$).

Given that both Allen et al. (2012) and Brownlees and Engle (2017) comprehensively investigated the ability of the CATFIN and SRISK measures, respectively, to forecast future macroeconomic conditions, we also consider forecasting models

[10] Data on the credit-to-GDP gap are quarterly, however, as suggested by an anonymous referee, we produce a monthly series by interpolating the quarterly values. This approach can be justified on the basis that the series is slow moving.

which include CATFIN or SRISK in place of our macro-financial spillover measures in the relevant predictive regressions[11]. In contrast to our macro-financial measures that incorporate information from the financial and real sides of the economy, both CATFIN and SRISK are more traditional financial systemic risk measures. The former measures the aggregate level of systemic risk in the financial system estimated via value-at-risk or expected shortfall, whereas SRISK utilises market and balance sheet data to estimate the expected capital shortfall of financial firms subject to the occurrence of a systemic event. Following the empirical analysis in Allen et al. (2012) and Brownlees and Engle (2017) respectively, CATFIN enters the predictive regressions in level form and SRISK as a log first difference or growth rate. It should be noted that data availability constraints for SRISK require us to start our sample period in June 2000, rather than January 1975 as in the previous section.

To examine the incremental value of our macro-financial spillover measures and the existing SRISK and CATFIN measures for forecasting macroeconomic conditions, we focus on evaluating forecast performance attainable when including these measures relative to an otherwise identical benchmark forecasting model that excludes the spillover or systemic risk measures. The form of the predictive regressions for the benchmark model is thus given by (7) with the $\beta_n s_t$ term excluded.

We have noted the challenges of model estimation and forecasting in the current economic environment due to the COVID-19 pandemic. Econometrically there is debate as to whether the extreme observations introduced by these events represent outliers, and as such, should either be down-weighted or even removed. For the plots of macro-financial spillover indexes in Section 3, we chose to include these pandemic observations in our sample period since they only affect the values of the spillover measures for those specific time periods and not those for the remainder of the sample period. In contrast, the formal tests of predictive ability we employ in the current section are based on the complete set of forecasts and will therefore be affected by this small number of extreme observations. We therefore evaluate predictive ability over two alternative sample periods ending in December 2019 and June 2020 respectively.

4.2. Evaluation of predictive ability

To formally evaluate forecast performance, we use the commonly employed test for equal predictive accuracy of Clark and West (2007), henceforth CW07, which provides a test of the null hypothesis that the forecasts obtained from two (possibly nested) forecasting models perform equally. In all cases, we compare the performance of forecasts produced by the models augmented with the various macro-financial spillover and financial systemic risk measures against those from the benchmark model described above. The rejection of the null hypothesis of equal predictive accuracy implies that the simpler benchmark model is outperformed by the extended model that incorporates spillover or systemic risk measures.

We begin with an in-sample forecasting environment, in which all forecasting models are estimated using data spanning the entire evaluation periods, and the predictive accuracy of the resulting in-sample forecasts over these periods is evaluated. Sample p-values for the CW07 test are reported in Table 1 for both our total and pairwise macro-financial spillover measures. We also include two combination forecasts obtained from the set of pairwise macro-financial spillover measures (columns 'PW mean' and 'PW med'), which are constructed as described below. Finally, we include forecasts that employ the existing SRISK and CATFIN systemic risk measures.

Beginning with the predictive accuracy of the six pairwise spillover measures in columns 2 to 7, we observe numerous increases in predictive accuracy over the benchmark model that are statistically significant across both sample periods. Although performance varies over forecasting horizons and across the various pairwise spillover measures, the number of statistically significant gains in predictive ability over the benchmark model for the pairwise spillover measures is marginally higher in panel (b) than the pre-pandemic sample in panel (a). Given that the CW07 test assesses predictive ability relative to the benchmark model (rather than in an absolute sense), this could be partly attributable to poorer performance of the benchmark model during the start of the pandemic.

The variation in predictive accuracy gains observed across spillover measures, sample periods, and forecasting horizons for the pairwise spillover measures suggests potential performance improvements can be obtained by combining the informational content of the individual spillover measures. One way to achieve this would be to extend the predictive regression in (7) to include multiple spillover measures simultaneously. However, it is frequently found in the literature that this approach of combining multiple predictors results in poorer performing forecasts than combining the forecasts obtained from distinct forecasting models containing different predictors. Examples from other areas of the finance literature include Rapach et al. (2010) who consider the problem of forecasting the equity premium and Paye (2012) who considers forecasts for equity market volatility using macroeconomic variables. The former also contains a concise discussion of the possible reasons for the solid empirical performance of combination forecasts.

The construction of combination forecasts is possible because the DY spillover approach naturally provides a set of directional pairwise measures, each measuring different aspects of the spillover structure. This contrasts with most existing systemic risk measures, which provide only higher-level summary measures in their standard form. Therefore we include two simple combination forecasts obtained as the mean and median of the forecasts[12] obtained from the six pairwise spillover measures, which are denoted by 'PW mean' and 'PW med' respectively in Table 1. Both of these combination forecasts perform

[11] Data for SRISK were kindly provided by the NYU V-Lab (https://vlab.stern.nyu.edu). Those for CATFIN were obtained from Turan Bali's website (https://sites.google.com/a/georgetown.edu/turan-bali).

[12] More complicated forecast combination methods are possible, such as those that weight the individual forecasts based on past performance. However, in practice, these are frequently found to perform similarly to the simple combinations such as the mean (see, e.g. Rapach et al., 2010).

Table 1
In-sample predictive accuracy for macroeconomic conditions.

Horizon (n)	Total	S&P to TNX	S&P to NAI	TNX to S&P	TNX to NAI	NAI to S&P	NAI to TNX	PW mean	PW med	SRISK	CAT- FIN
(a) Sample period 2000:06–2019:12											
n = 1	0.279	0.147	0.470	0.133	0.489	0.030	0.039	0.002	0.012	0.151	0.268
n = 2	0.241	0.262	0.145	0.204	0.439	0.020	0.044	0.002	0.008	0.323	0.024
n = 3	0.123	0.098	0.153	0.046	0.408	0.049	0.028	0.005	0.013	0.401	0.010
n = 4	0.173	0.118	0.258	0.052	0.355	0.081	0.027	0.011	0.009	0.189	0.017
n = 5	0.165	0.160	0.166	0.080	0.377	0.127	0.036	0.027	0.024	0.192	0.008
n = 6	0.123	0.120	0.124	0.057	0.413	0.208	0.028	0.027	0.029	0.403	0.041
n = 7	0.130	0.133	0.118	0.062	0.314	0.463	0.046	0.053	0.064	0.180	0.014
n = 8	0.151	0.128	0.106	0.074	0.151	0.346	0.040	0.048	0.056	0.178	0.027
n = 9	0.183	0.133	0.120	0.076	0.071	0.298	0.036	0.040	0.045	0.137	0.005
n = 10	0.269	0.179	0.149	0.123	0.038	0.210	0.033	0.039	0.050	0.047	0.000
n = 11	0.341	0.189	0.101	0.160	0.019	0.325	0.027	0.029	0.033	0.016	0.002
n = 12	0.404	0.220	0.090	0.221	0.016	0.245	0.023	0.027	0.034	0.116	0.028
(b) Sample period 2000:06–2020:06											
n = 1	0.108	0.126	0.109	0.157	0.106	0.060	0.082	0.095	0.087	0.103	0.089
n = 2	0.483	0.427	0.059	0.266	0.252	0.410	0.132	0.129	0.169	0.102	0.153
n = 3	0.039	0.039	0.101	0.057	0.194	0.062	0.063	0.057	0.069	0.148	0.118
n = 4	0.033	0.014	0.108	0.026	0.229	0.251	0.044	0.044	0.043	0.155	0.315
n = 5	0.037	0.026	0.111	0.009	0.226	0.439	0.475	0.066	0.097	0.153	0.406
n = 6	0.016	0.008	0.098	0.003	0.225	0.358	0.487	0.057	0.080	0.202	0.410
n = 7	0.010	0.011	0.090	0.005	0.203	0.408	0.310	0.053	0.075	0.046	0.370
n = 8	0.008	0.008	0.080	0.003	0.222	0.427	0.261	0.021	0.032	0.126	0.086
n = 9	0.012	0.007	0.077	0.004	0.353	0.344	0.223	0.012	0.030	0.422	0.046
n = 10	0.013	0.006	0.076	0.007	0.169	0.204	0.171	0.018	0.046	0.074	0.149
n = 11	0.011	0.008	0.071	0.004	0.130	0.135	0.474	0.038	0.048	0.210	0.055
n = 12	0.026	0.007	0.072	0.014	0.116	0.219	0.231	0.019	0.033	0.065	0.326

The table reports sample p-values for the CW07 test of equal predictive accuracy applied to in-sample n-step-ahead forecasts for the level of the CFNAI over the sample periods 2000:06–2019:12 (panel (a)) and 2000:06–2020:06 (panel (b)). Forecasts are obtained using predictive regressions of the form given in Eq. (7) containing the current value of a single spillover or systemic risk measure, with the exception of the combination forecasts. For pairwise spillovers, equities, bonds and the real economy are denoted by S&P500, TNX and NAI respectively. The combination forecasts are labelled as 'PW mean' and 'PW med' and are constructed respectively as the mean and median of the forecasts obtained from the 6 pairwise spillover measures. The null hypothesis is that the forecasts for the relevant model and the benchmark model have equal predictive accuracy. Rejection of the null implies that the extended model has superior predictive accuracy to the benchmark model.

strongly, providing improvements in predictive accuracy over the benchmark model that are statistically significant at the 10% level in all but 2 cases (both in panel (b)), and significant at either the 5% or 1% levels in the vast majority of cases. Inclusion of the COVID-19 pandemic slightly reduces the statistical significance of the gains in predictive accuracy relative to the results in panel (a), however the performance of the combination forecasts is still solid. Given their consistently strong performance, the ability to easily construct these combination forecasts is a significant advantage of the DY approach in this context.

Turning next to forecasts based on the total macro-financial spillover index, it can be seen that the overall level of macro-financial spillovers has relatively poor predictive accuracy for macroeconomic conditions in the in-sample context in panel (a), failing to provide gains in predictive accuracy over the benchmark model that are statistically significant at the 10% level at any of the forecast horizons. In panel (b) on the other hand, statistically significant gains are observed at all forecast horizons longer than 2 months. As noted above, this may be attributable to a large drop in the predictive ability of the benchmark model during the first half of 2020.

This lack of consistency in the forecasting gains from the total index relative to the individual pairwise measures indicates that whilst the total spillover index provides a useful summary measure that is conceptually more similar to existing measures in the literature, the ability to decompose it into pairwise spillovers and construct combination forecasts is a valuable benefit of our DY-based approach in this context.

Finally, moving on to the predictive ability of the existing SRISK and CATFIN measures, the former fails to produce statistically significant gains over the benchmark model except for a small number of longer forecast horizons. The CATFIN measure, on the other hand, provides improvements in predictive accuracy across all forecast horizons except the shortest one-month horizon in panel (a) that are significant at either the 5% or 1% levels. When the first half of 2020 is added in panel (b) however, the relative predictive ability of CATFIN drops substantially, only providing statistically significant gains at horizons of 8, 9 and 11 months.

We next evaluate the out-of-sample forecasting performance of the various models relative to the same benchmark model used above. In all cases, out-of-sample forecasts are produced using a standard rolling-window approach with a fixed window length of 60 months. Data within the window are used to estimate the parameters of the relevant predictive regression and produce the n-step-ahead forecast, with the model parameters re-estimated each time the window is rolled forward. The out-of-sample combination forecasts are computed as the mean or median of the out-of-sample forecasts obtained from the pairwise spillover measures. Such a pseudo-out-of-sample forecasting environment arguably better rep-

resents how these measures would potentially be used in practice by an individual attempting to forecast future economic conditions in real time and, as argued by Diebold (2015), can be used to assess forecasting performance during different historical periods.

To evaluate out-of-sample forecasting performance, we again employ the CW07 test for equal predictive accuracy but supplement this with the out-of-sample R^2 measure of Campbell and Thompson (2008). For a series of out-of-sample forecasts produced for periods $t = 1, \ldots, S$, the out-of-sample R^2 measure is computed as:

$$R^2_{OS} = 1 - \frac{\sum_{t=1}^{S} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{S} (y_t - \hat{y}_t^b)^2}$$

where y_t is the actual value of the series to be predicted, and $\hat{y}_t$ and $\hat{y}_t^b$ are the forecasted values from the model under consideration and the benchmark model respectively. As such, positive (negative) values of R^2_{OS} imply that the forecasting model under consideration has a lower (higher) mean-squared prediction error (MSPE) than the benchmark model.

Out-of-sample forecasting results are presented in Table 2, with those for the 2000:06–2019:12 sample period in Panel (a) and the 2000:06–2020:06 sample in Panel (b). To conserve space, we exclude results for the predictive regressions containing each of the individual pairwise macro-financial spillover measures[13] and include only the combination forecasts obtained from these pairwise forecasts.

Beginning with the Campbell and Thompson (2008) out-of-sample R^2 values, in most cases, the forecasts obtained using macro-financial spillover measures result in lower mean squared prediction errors than the benchmark model. Most notably, the strong performance of the combination forecasts from the pairwise spillovers continues in the out-of-sample case, dominating the benchmark model at all but the shortest horizons and displaying larger out-of-sample R^2 values than those for the total spillover index in almost all cases. These predictive gains relative to the benchmark model are typically larger at longer forecast horizons.

In contrast, the predictive performance of SRISK and CATFIN is mixed. Forecasts incorporating SRISK often have a larger MSPE than the benchmark model for forecast horizons of less than five months but typically outperform it for longer horizons. In cases where the SRISK-based forecasts result in a lower MSPE than the benchmark, the out-of-sample R^2 values obtained are lower than those for the pairwise combination forecasts in every case. Furthermore, whilst CATFIN exhibited strong forecasting power in-sample in Table 1, this largely disappears when we look at the out-of-sample results. The forecasts incorporating CATFIN result in out-of-sample R^2 values that are negative or close to zero in almost every case.

The out-of-sample results for the CW07 test reinforce the strong in-sample predictive accuracy of the pairwise mean and median combination spillover forecasts, with gains over the benchmark model that are statistically significant in almost all cases, with the main exceptions being the shortest 1- to 3-month horizons. Interestingly, compared to the previous in-sample forecasting environment, the out-of-sample forecasting performance of the total macro-financial spillover index relative to the benchmark model improves substantially, producing statistically significant gains at the 5% level for almost all forecast horizons longer than three months. Indeed, forecasts utilising the total spillover index display the lowest p-values of all methods for horizons of 5 to 12 months, including the two combination forecasts.

Turning finally to the forecasts obtained using the SRISK and CATFIN measures, the CW07 test results largely confirm the general patterns discussed previously in the context of the out-of-sample R^2 values. SRISK exhibits statistically significant gains in predictive ability over the benchmark model at many of the longer forecast horizons, being broadly competitive with the pairwise combination forecasts, but falling behind forecasts employing the total spillover index. The CATFIN-based forecasts, on the other hand, only provide statistically significant gains at the 10% level at a small number of horizons for the pre-pandemic sample in panel (a) and at no horizons in panel (b). As previously noted for the out-of-sample R^2, this is in contrast to the previous in-sample environment where highly statistically significant improvements over the benchmark model were observed at almost all horizons.

In a similar manner to Giglio et al. (2016) we next investigate the ability of the various spillover and systemic risk measures to forecast the four disaggregated subcomponents of the CFNAI index, which consist of production and income (PI), employment, unemployment and hours (EUH), consumption and housing (CH), and sales, orders and inventories (SOI). This allows us to investigate whether the relative predictive ability of the measures varies according to the specific aspects of macroeconomic conditions that are being forecasted. Table 3 presents p-values for the CW07 test for both in-sample and out-of-sample forecasts constructed using the same approach as employed above. To conserve space only results for a subset of the previous forecasting horizons are presented.

Beginning with the in-sample results, the solid predictive accuracy of the pairwise spillover combination forecasts for the aggregate CFNAI index in Table 1 carries through for the subcomponents of the index, with highly statistically significant gains over the benchmark in all cases. For the total macro-financial spillover index, forecasting performance depends strongly on the subcomponent to be forecasted, with predictive accuracy consistently strong when forecasting the EUH subcomponent and to a lesser extent the SOI subcomponent, but more variable across forecast horizons and weaker on average

[13] These results are broadly similar to the previous in-sample case, with the pairwise measures providing statistically significant gains in predictive accuracy over the benchmark model in many cases, but performance varying across horizons, spillover measures and sample periods.

Table 2
Out-of-sample R^2 values and tests of equal predictive accuracy.

	Total		PW mean		PW median		SRISK		CATFIN	
Horizon (n)	R^2	CW	R^2	CW	R^2	CW	R^2	CW	R^2	CW
					(a) Sample period 2000:06–2019:12					
n = 1	-0.074	0.790	-0.007	0.393	0.003	0.164	-0.016	0.650	0.003	0.145
n = 2	-0.027	0.205	0.024	0.039	0.002	0.199	-0.017	0.665	0.005	0.093
n = 3	0.103	0.030	0.145	0.024	0.104	0.042	-0.041	0.929	0.031	0.048
n = 4	0.093	0.020	0.141	0.018	0.108	0.023	-0.006	0.239	0.033	0.075
n = 5	0.141	0.011	0.159	0.019	0.136	0.024	0.005	0.184	0.000	0.208
n = 6	0.127	0.007	0.151	0.015	0.136	0.023	0.014	0.030	-0.005	0.253
n = 7	0.153	0.010	0.216	0.027	0.223	0.050	0.011	0.094	0.007	0.196
n = 8	0.163	0.011	0.248	0.034	0.247	0.060	0.009	0.074	-0.018	0.467
n = 9	0.100	0.005	0.183	0.024	0.182	0.047	0.035	0.032	0.068	0.093
n = 10	0.086	0.001	0.222	0.023	0.193	0.036	0.118	0.013	0.012	0.142
n = 11	0.077	0.001	0.246	0.014	0.199	0.021	0.097	0.045	0.012	0.087
n = 12	0.058	0.006	0.242	0.021	0.195	0.017	0.070	0.067	-0.009	0.236
					(b) Sample period 2000:06–2020:06					
n = 1	-0.151	0.840	-0.096	0.852	-0.026	0.861	0.128	0.158	0.281	0.160
n = 2	-0.139	0.837	-0.004	0.719	0.035	0.151	0.018	0.161	0.059	0.150
n = 3	-0.017	0.415	0.005	0.226	0.007	0.081	-0.004	0.962	0.000	0.122
n = 4	0.009	0.032	0.019	0.024	0.013	0.038	-0.011	0.809	-0.001	0.215
n = 5	0.021	0.012	0.026	0.017	0.020	0.030	0.004	0.119	0.004	0.120
n = 6	0.017	0.014	0.025	0.026	0.018	0.058	0.002	0.065	-0.002	0.273
n = 7	0.022	0.017	0.032	0.066	0.029	0.124	0.003	0.052	-0.004	0.383
n = 8	0.031	0.013	0.046	0.044	0.049	0.066	0.002	0.061	0.015	0.157
n = 9	0.034	0.002	0.043	0.024	0.049	0.034	0.008	0.039	0.002	0.232
n = 10	0.017	0.002	0.052	0.018	0.041	0.039	0.014	0.110	0.000	0.225
n = 11	0.010	0.005	0.053	0.012	0.038	0.025	0.019	0.050	-0.001	0.146
n = 12	0.000	0.091	0.044	0.053	0.039	0.035	0.022	0.038	-0.003	0.283

The table reports Campbell and Thompson (2008) out-of-sample R^2 values and sample p-values for the CW07 test of equal predictive accuracy from predictive regressions for out-of-sample n-step-ahead forecasts for the level of the CFNAI over the sample periods 2000:06–2019:12 (panel (a)) and 2000:06–2020:06 (panel (b)). Forecasts are obtained using predictive regressions of the form given in Eq. (7) containing the current value of a single spillover or systemic risk measure, with the exception of the combination forecasts. The combination forecasts are labelled 'PW mean' and 'PW med' and are constructed respectively as the mean and median of the forecasts obtained from the 6 pairwise spillover measures. Out-of-sample forecasts for all models are obtained using a standard rolling window approach with a window length of 60 months. The Campbell and Thompson (2008) out-of-sample R-squared values are computed for each model relative to the benchmark model that excludes spillover or systemic risk measures. For the Clark and West (2007) test the null hypothesis is that the forecasts for the relevant model and the benchmark model have equal predictive accuracy. Rejection of the null implies that the extended model has superior predictive accuracy to the benchmark model.

for the remaining subcomponents. We hypothesise that this arises because of differences in the responsiveness of the underlying macroeconomic series to changes in the specific dimension of current economic conditions that the total spillover index proxies. For example, series within the consumption and housing subcomponent are likely to be sluggish and slower to respond to changes in macro-financial conditions than sales or orders in the SOI subcomponent, or employee hours in the EUH subcomponent. The same logic applies to the PI component, which is comprised primarily of industrial production series, in addition to construction and personal income. This suggests that in the case of some predictors, the relative gains in predictive accuracy do indeed depend on the specific aspect of macroeconomic conditions to be forecasted.

Both SRISK and CATFIN fail to exhibit consistent in-sample predictive power for the CFNAI subcomponents, providing statistically significant improvements over the benchmark approach in only a small number of cases across the various forecast horizons and index subcomponents. For the case of CATFIN, this is limited to the 1- to 3-month horizons and contrasts sharply to its consistently strong in-sample performance in Table 1 when forecasting the aggregate CFNAI index. This is particularly notable given that the pairwise spillover combination forecasts manage to preserve their previous performance for the case of the CFNAI subcomponents.

Turning to out-of-sample forecasting performance, there is strong evidence of gains in predictive accuracy for the four CFNAI subcomponents using either the total spillover index or the two pairwise spillover forecast combination schemes. The various forecasts incorporating macro-financial spillover measures typically fail to provide statistically significant gains in predictive accuracy over the benchmark model only at the shortest one and 2-month forecast horizons. As with the in-sample case, CATFIN and SRISK only provide statistically significant improvements in out-of-sample predictive accuracy over the benchmark model for the CFNAI in a small number of cases. For the SRISK-based forecasts, gains are observed only for longer forecast horizons for the PI and EUH subcomponents.

Table 3
In-sample and out-of-sample predictive accuracy for alternative measures of macroeconomic conditions.

	In-sample					Out-of-sample				
Horizon (n)	Total	PW mean	PW med	SRISK	CAT- FIN	Total	PW mean	PW med	SRISK	CAT- FIN
					(a) Production and income (PI)					
n = 1	0.355	0.004	0.017	0.105	0.071	0.810	0.158	0.017	0.365	0.122
n = 2	0.319	0.009	0.049	0.390	0.043	0.570	0.334	0.542	0.387	0.161
n = 3	0.148	0.018	0.051	0.466	0.047	0.045	0.026	0.062	0.937	0.080
n = 6	0.091	0.043	0.038	0.379	0.321	0.009	0.006	0.017	0.018	0.230
n = 9	0.173	0.088	0.118	0.167	0.031	0.008	0.009	0.032	0.058	0.107
n = 12	0.479	0.072	0.097	0.098	0.257	0.047	0.012	0.008	0.087	0.177
					(b) Employment, unemployment and hours (EUH)					
n = 1	0.301	0.000	0.009	0.115	0.203	0.811	0.008	0.119	0.331	0.046
n = 2	0.058	0.001	0.005	0.188	0.046	0.244	0.009	0.082	0.904	0.270
n = 3	0.019	0.001	0.003	0.418	0.204	0.017	0.005	0.012	0.368	0.151
n = 6	0.007	0.002	0.002	0.372	0.108	0.010	0.012	0.023	0.958	0.131
n = 9	0.009	0.004	0.011	0.188	0.052	0.001	0.007	0.021	0.044	0.104
n = 12	0.010	0.007	0.041	0.085	0.061	0.000	0.012	0.015	0.036	0.133
					(c) Consumption and housing (CH)					
n = 1	0.020	0.000	0.001	0.338	0.220	0.551	0.059	0.321	0.472	0.233
n = 2	0.222	0.000	0.000	0.442	0.058	0.552	0.029	0.042	0.726	0.152
n = 3	0.465	0.001	0.003	0.500	0.021	0.025	0.009	0.014	0.955	0.022
n = 6	0.330	0.000	0.000	0.490	0.192	0.004	0.002	0.007	0.861	0.221
n = 9	0.343	0.001	0.001	0.150	0.080	0.005	0.005	0.007	0.806	0.106
n = 12	0.472	0.001	0.001	0.441	0.335	0.056	0.006	0.004	0.637	0.057
					(d) Sales, orders and inventories (SOI)					
n = 1	0.177	0.003	0.040	0.484	0.151	0.721	0.343	0.473	0.973	0.175
n = 2	0.058	0.006	0.043	0.013	0.151	0.215	0.163	0.160	0.528	0.127
n = 3	0.048	0.010	0.018	0.268	0.013	0.008	0.012	0.042	0.921	0.198
n = 6	0.120	0.046	0.028	0.395	0.047	0.014	0.016	0.034	0.974	0.258
n = 9	0.162	0.072	0.057	0.227	0.168	0.030	0.027	0.040	0.580	0.280
n = 12	0.424	0.038	0.027	0.191	0.419	0.057	0.028	0.024	0.138	0.349

The table reports sample p-values for the CW07 test of equal predictive accuracy applied to in-sample and out-of-sample n-step-ahead forecasts for the various subcomponents of the CFNAI index over the period 2000:06–2019:12. CFNAI subcomponents are production and income (PI), employment, unemployment and hours (EUH), consumption and housing (CH), and sales, orders and inventories (SOI). Forecasts are obtained using predictive regressions of the form given in Eq. (7) containing the current value of a single spillover or systemic risk measure, with the exception of the combination forecasts. The combination forecasts are labelled 'PW mean' and 'PW med' and are constructed respectively as the mean and median of the forecasts obtained from the 6 pairwise spillover measures. Out-of-sample forecasts for all models are obtained using a standard rolling window approach with a window length of 60 months. The null hypothesis for the Clark and West (2007) test in each case is that the forecasts for the relevant model and the benchmark model have equal predictive accuracy, with rejection of the null implying that the extended model has superior predictive accuracy to the benchmark model.

4.3. Time-variation in predictive accuracy

Whilst the CW07 test and the Campbell and Thompson (2008) out-of-sample R^2 measures are both employed widely in the literature to compare predictive accuracy between competing forecasting models; they do not directly reveal anything about how the relative performance of the alternative models may vary over the forecast evaluation period. In particular, it may be the case that specific predictors of future macroeconomic conditions have a greater predictive ability during some time periods than others, depending on the state of the economy.

To shed light on this and other related issues, we construct plots of the cumulative squared prediction errors (CSPEs) from predictive regressions for the n-period-ahead values of the CFNAI relative to that of the benchmark model. Such plots are a standard tool in the forecasting literature, with two examples including Rapach et al. (2010) and Paye (2012). We subtract the values of the cumulative squared prediction errors obtained for the predictive regression of interest over the in-sample or out-of-sample period from those obtained from the previous benchmark model that excludes all spillover and systemic risk measures. Values larger (smaller) than zero imply that the forecasting model in question has a lower (higher) cumulative squared prediction error than the benchmark model up to that point in time. An increase (decrease) during a specific time period implies that the relevant model is currently outperforming (underperforming) the benchmark model.

To conserve space, we only consider a subset of the forecasting horizons and include results for the total spillover index, the two combination forecasts constructed from the pairwise spillover forecasts, SRISK and CATFIN. Whilst we use the same forecast evaluation periods as before for consistency, it should be noted that the vertical position or height of each line at a given point in time is affected by the chosen start date for each plot (given that all must equal zero at the start of the chosen evaluation period). However, the shape or gradient of each line is unaffected by this choice.

Fig. 5 plots the relative CSPEs for the previous 2000:06–2019:12 sample period. We observe strong performance for our pairwise spillover combination forecasts, particularly during the 2008–2009 crisis. At the 1 and 2 month horizons, the forecasting performance of the total spillover index remains close to that of the benchmark model throughout the evaluation period, deviating only slightly from zero. For horizons of greater than three months, forecasts obtained from the total spil-

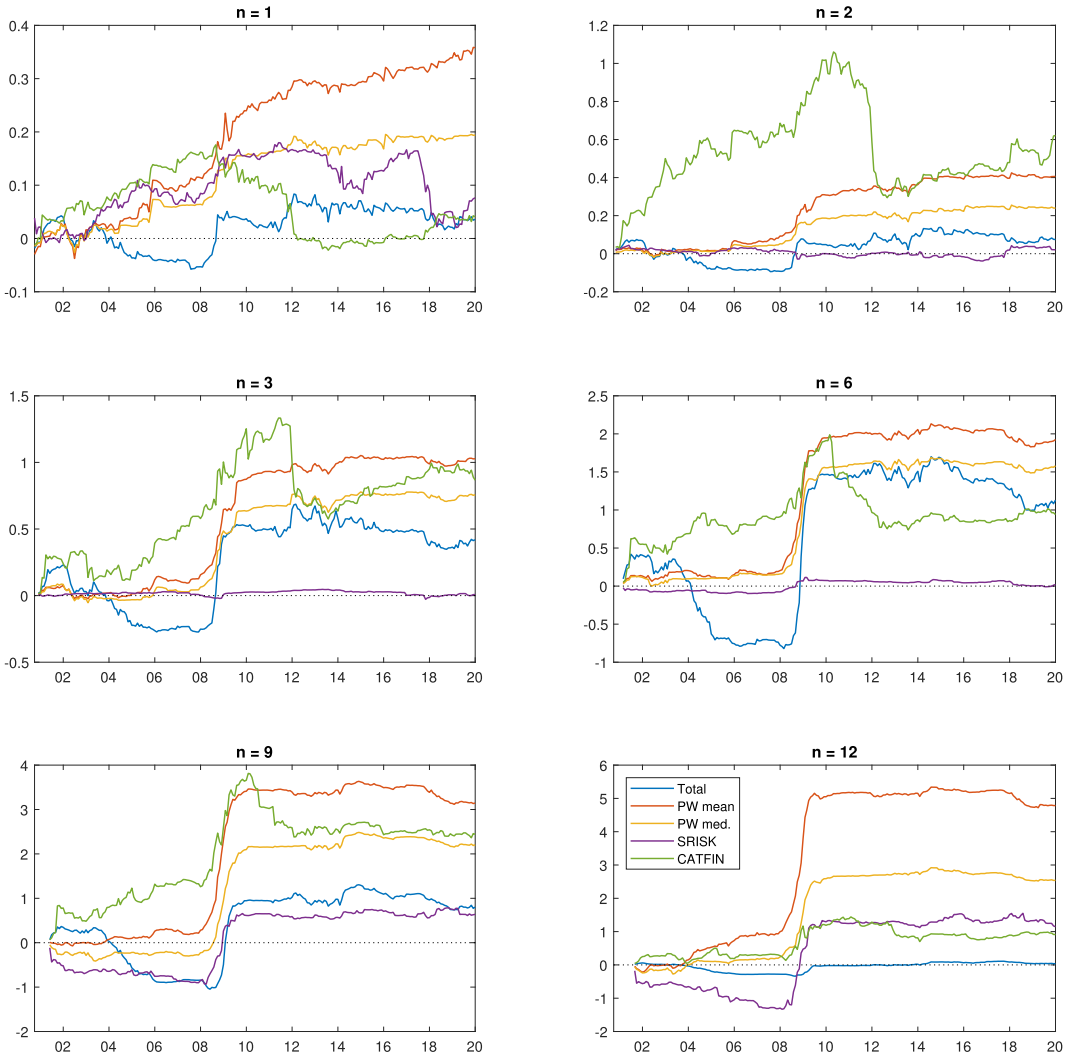

Fig. 5. In-sample cumulative squared prediction errors relative to the benchmark model. The figure plots cumulative squared in-sample prediction errors of the benchmark predictive regression model minus the cumulative squared prediction errors of the predictive regression models that include the total macro-financial spillover index, the combination forecasts computed as the mean and median of the 6 individual pairwise spillover index forecasts, SRISK or CATFIN. Larger values correspond to stronger performance of the relevant extended predictive regression model relative to the benchmark, with values above (below) zero implying a smaller (larger) cumulative squared prediction error than the benchmark model up to that point in time. The sample period considered is 2000:06–2019:12 and the dependent variable to be forecasted in all cases is the *n*-period-ahead level of the CFNAI..

lover index perform worse than the benchmark model in the early parts of the evaluation period, with the relative CSFE values becoming steadily more negative until around 2006. However, after the start of the financial crisis in 2008, there was a large and sharp increase in the relative CSFE until the peak of the crisis had subsided in around 2010, suggesting substantial gains in predictive accuracy for the total spillover index during the crisis. Even more substantial gains in predictive accuracy are typically observed for the pairwise mean combination forecasts during the crisis period at the 3 to 12-month forecast horizons. However, unlike the forecasts obtained from the total spillover index, they typically outperform the benchmark over the pre-crisis period, too, resulting in substantially higher relative CSFEs when considered over the complete evaluation period. Finally, the forecasts obtained from the pairwise median combination forecast typically lie between the total spillover index and the mean combination forecast.

The performance of the SRISK-based forecasts is nearly identical to that of the benchmark model throughout the 2000:06–2019:12 evaluation period for the intermediate 2, 3 and 6-month forecast horizons. At the longer nine and 12-month horizons, the dynamics of forecast performance are somewhat similar to those of the forecasts based on the total spillover index, with performance inferior to the benchmark in the pre-crisis period, superior to it during the crisis and comparable to it in the post-crisis period. Considering the CATFIN-based forecasts finally, we generally observe strong performance at the start of the evaluation period followed by a further increase during the crisis period. However, this peak coinciding

with the 2008 crisis is substantially less pronounced than that exhibited by the forecasts employing macro-financial spillover measures and is typically followed by a decrease in performance relative to the benchmark model.

Fig. 6 presents equivalent plots to Fig. 5 for the out-of-sample case, with all out-of-sample forecasts produced using the rolling window approach discussed above. Given the rolling window approach used to produce out-of-sample forecasts and the lead and lag values contained in Eq. (7), the forecast evaluation periods are shorter than the in-sample case and vary with the forecast horizon.

In the out-of-sample forecasting environment, the differences in cumulative squared forecast errors for the augmented models compared to the benchmark model tend to be relatively small at the beginning of the evaluation period and also from 2012–2014 (depending on the forecast horizon) until the start of the COVID-19 pandemic, as indicated by the relatively flat CSPE plots during these periods. Indeed, the most significant differences in predictive accuracy amongst the various forecasting models and the benchmark model occur during and immediately following the 2008 financial crisis and at the start of the ongoing COVID-19 pandemic.

In most cases, the forecasts based on the total spillover index, the pairwise forecast combinations and those based on CATFIN again exhibit apparent increases in predictive accuracy relative to the benchmark model during and following the 2008 crisis, though the size and duration of these increases vary substantially across forecast horizons. The main exceptions are forecasts based on the total spillover index at the shortest 1-month horizon and those based on CATFIN at the longest 12-month horizon. Both perform poorly during this period. In the case of the pairwise combination forecasts, these gains tend to be substantial around the crisis and continue to accumulate gradually even after the financial crisis has passed and are maintained until the end of the sample period. By contrast, in the case of forecasts based on the total spillover index and CATFIN, some or all of the gains in cumulative forecast accuracy over the benchmark model attained during the crisis are generally lost in the post-crisis period, with overall cumulative performance falling below that of the benchmark model in some cases from 2011 onwards. Finally, the SRISK-based forecasts exhibit gains during the 2008 crisis only for the longer nine and 12-month forecast horizons.

5. Conclusion

We estimate and analyse the structure of macro-financial spillovers between equities, bonds and the real side of the economy. For this purpose, we develop a new approach for estimating macro-financial spillovers that combines established quantitative measures of financial spillovers with mixed-frequency econometric methods. Our approach permits the use of mixed-frequency macro-financial datasets without the need to aggregate the higher frequency financial series down to the lower frequency as the real economy series. This is in contrast to existing measures of spillovers and systemic risk previously proposed in the literature, which are purely financial in nature and so cannot directly estimate connectedness between the real and financial sides of the economy. Furthermore, our methodology produces a set of different macro-financial spillover measures that consider the direction of spillovers. The directionality of the measures obtained permits a more detailed analysis of market linkages than other existing approaches that measure only the strength of association.

In our analysis of macro-financial spillovers in the US economy from 1975 to 2020, we find that the magnitude of the mixed-frequency spillovers is substantially greater than those obtained from a similar common-frequency approach. This suggests that the loss of high-frequency information incurred by using a common-frequency modelling approach results in the financial and real sides of the economy appearing less connected. Furthermore, our mixed-frequency approach's preservation of additional high-frequency information results in spillover measures that appear more consistent with key events that occurred during our sample period. The same empirical findings are also evident in a set of two-country analyses of international macro-financial spillovers. The directional nature of our spillover measures demonstrates that the largest magnitude of spillovers originates from the financial, rather than the real, side of the economy. This decomposition also clearly shows that the relative importance of each financial market has changed over time.

Motivated by existing work such as Allen et al. (2012) and Brownlees and Engle (2017) that analyses the predictive ability of financial systemic risk measures for future macroeconomic series, we explore whether our macro-financial spillover measures can be employed to forecast US macroeconomic conditions. Consistent with the financial measures in these existing studies, we find that forecasts produced using our macro-financial spillover measures predict both broad measures of overall macroeconomic conditions and also measures representing more specific aspects of the state of the macroeconomy.

Forecasting performance is particularly strong for the case of simple combination forecasts computed as the mean or median of the individual forecasts obtained from our set of pairwise macro-financial spillover measures. In an in-sample forecasting environment, these combination forecasts marginally outperform existing financial systemic risk measures when forecasting aggregate macroeconomic conditions. However, they substantially outperform them when forecasting disaggregated measures representing more specific aspects of economic conditions. When moving to an out-of-sample forecasting environment, the gains in the predictive accuracy of our macro-financial spillover measures are typically even larger when forecasting either aggregate or more specific aspects of macroeconomic conditions. When examining the dynamics of forecasting performance, we find that the gains in predictive accuracy provided by our spillover measures are especially large during the 2008–2009 crisis, both relative to our chosen benchmark model and also relative to those provided by existing financial systemic risk measures.

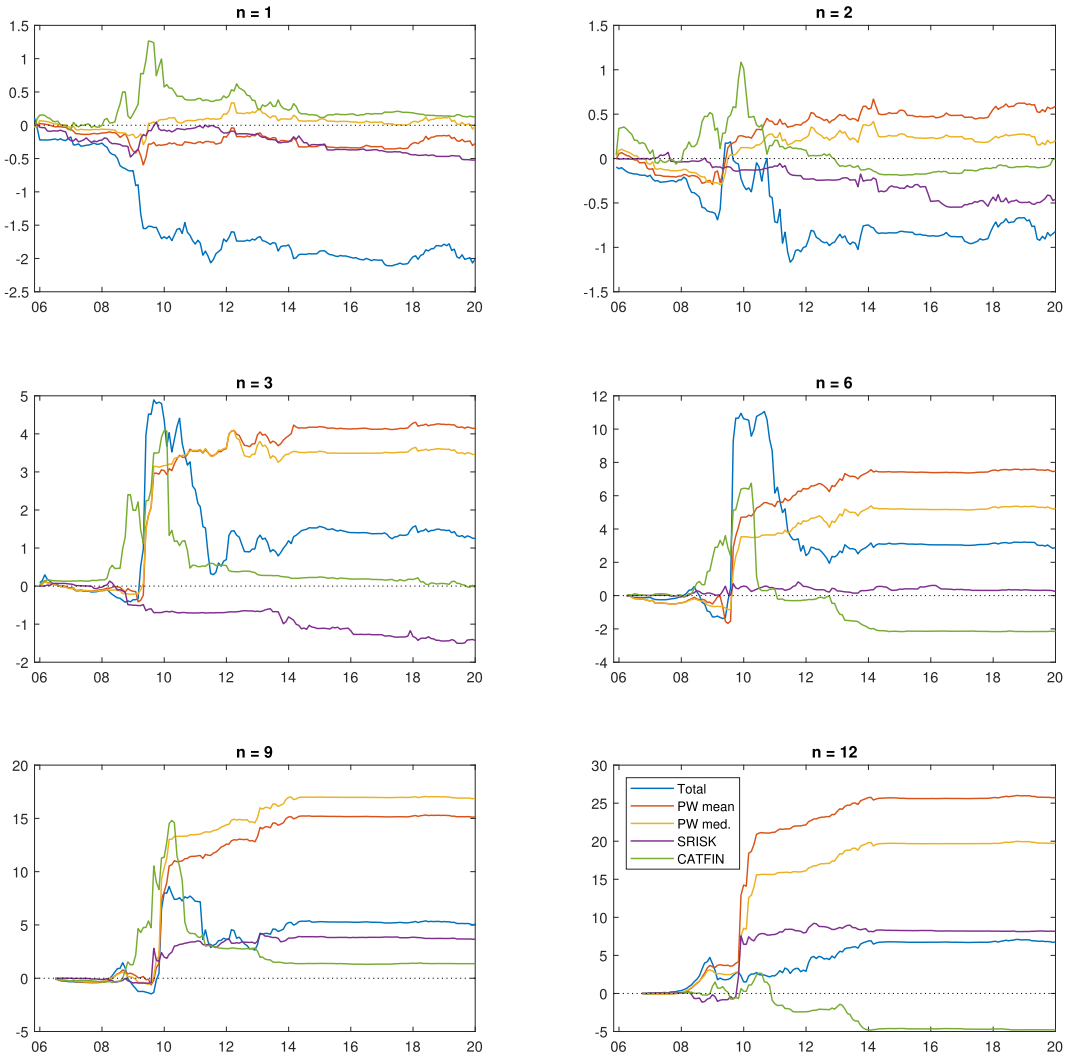

Fig. 6. Out-of-sample cumulative squared forecast errors relative to the benchmark model. The figure plots cumulative squared out-of-sample forecast errors of the benchmark predictive regression model minus the cumulative squared forecast errors of the predictive regression models that include the total macro-financial spillover index, the combination forecast computed as the mean and median of the pairwise spillover index forecasts, SRISK or CATFIN. Larger values correspond to the stronger performance of the relevant extended predictive regression model relative to the benchmark, with values above (below) zero implying a smaller (larger) cumulative squared forecast error than the benchmark model. The dependent variable to be forecasted in all cases is the *n*-period-ahead level of the CFNAI. Forecasts are produced using a rolling window scheme with a fixed window length of 60 months using data spanning the period 2000:06–2019:12. Forecasts are produced up to 2019:12 in all cases, with the date of the first forecast to be evaluated varying from 2005:10 to 2006:09 depending on the forecast horizon..

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Forecast error variance decomposition

A.1. Generalised forecast error variance decomposition

We denote the generalised FEVD values by $\theta_{ij}(H)$, where $\theta_{ij}(H)$ measures the fraction of the total H-step-ahead error variance in forecasting series i attributable to shocks in series j. Following Pesaran and Shin (1998), the generalised forecast error variance decomposition values are computed for any given forecast horizon $H = 1, 2, \ldots$ as:

$$\theta_{ij}(H) = \frac{\sigma_{jj}\sum_{h=0}^{H-1}\left(e_i'B_h\Sigma e_j\right)^2}{\sum_{h=0}^{H-1}\left(e_i'B_h\Sigma B_h'e_i\right)} \quad \text{for} i,j = 1,\ldots,K_x \tag{A.1}$$

where Σ is the covariance matrix of the error vector $\underline{\varepsilon}(\tau_L)$, σ_{jj} is the j-th diagonal element of Σ and e_j is the K_x-dimensional selection vector with a 1 in the j-th element and zeros elsewhere. The arrays $B_i, i = 1,\ldots$ are the coefficient arrays from the infinite order moving average (MA) representation of the MF-VAR in Eq. (1). These can be obtained from the coefficient arrays of the standard representation of the VAR via a simple recursion (see Diebold and Yılmaz, 2014 for details).

It is worth noting that unlike other common approaches to computing the FEVD that rely on orthogonalisation to account for potential correlation between shocks, such as the Cholesky decomposition, the values of the generalised FEVD arrays are not affected by the ordering of the series within the VAR. Instead, the approach accounts for potential correlation between shocks using the historical distribution of the errors.

A.2. Transformation of MF-VAR FEVD Arrays

We continue to employ the simple bivariate example from Section 2.3 for illustration, repeating some of the key details here for convenience. With one low-frequency monthly series and one high-frequency weekly series we obtain a (5×1) vector process with the form $\underline{x}(\tau_L) = [x_H(\tau_L,1),\ldots,x_H(\tau_L,4),x_L(\tau_L)]'$ for the MF-VAR, and a (2×1) vector process $\overline{x}(\tau_L) = [x_{HtL}(\tau_L),x_L(\tau_L)]'$ for the corresponding CF-VAR. This results in (5×5) and (2×2) FEVD arrays, given respectively by:

$$\begin{bmatrix} \theta_{11}(H) & \ldots & \theta_{15}(H) \\ \vdots & \ddots & \vdots \\ \theta_{51}(H) & \ldots & \theta_{55}(H) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \phi_{11}(H) & \phi_{12}(H) \\ \phi_{21}(H) & \phi_{22}(H) \end{bmatrix} \quad \text{for} \quad H = 1,2,\ldots \tag{A.2}$$

We argued previously that the elements of the MF-VAR FEVD arrays can be grouped into sub-arrays as:

$$\begin{bmatrix} \Theta_{11}(H) & \Theta_{12}(H) \\ \Theta_{21}(H) & \Theta_{22}(H) \end{bmatrix} \quad \text{for} \quad H = 1,2,\ldots \tag{A.3}$$

where:

$$\Theta_{11}(H) \equiv \begin{bmatrix} \theta_{11}(H) & \ldots & \theta_{14}(H) \\ \vdots & \ddots & \vdots \\ \theta_{41}(H) & \ldots & \theta_{44}(H) \end{bmatrix} \qquad \Theta_{12}(H) \equiv \begin{bmatrix} \theta_{15}(H) \\ \vdots \\ \theta_{45}(H) \end{bmatrix}$$

$$\Theta_{21}(H) \equiv [\,\theta_{51}(H) \quad \ldots \quad \theta_{54}(H)\,] \qquad \Theta_{22}(H) \equiv \theta_{55}(H)$$

such that each of the sub-arrays $\Theta_{kl}(H)$ in (6) can be viewed as a mixed-frequency analogue of the corresponding scalar element $\phi_{kl}(H)$ from the CF-VAR FEVD array. Our transformation approach produces new FEVD arrays from the MF-VAR FEVD arrays with the same structure and dimensions as those for the corresponding CF-VAR. We denote a generic element of the new transformed FEVD arrays by $\psi_{kl}(H)$ for $k,l = 1,\ldots,K$.

The key is to perform the transformation such that the value and interpretation of each element $\psi_{kl}(H)$ is directly comparable with the corresponding element $\phi_{kl}(H)$. This relies on the correspondence between the elements of the mixed and common frequency arrays discussed above and the mathematical definition of the generalised FEVD elements in Eq. (A.1). For ease of notation, we denote the numerator and denominator of (A.1) more compactly as:

$$\theta_{ij}(H) = \frac{\lambda_{ij}(H)}{\mu_i(H)} \quad \text{for} i,j = 1,\ldots,K_x \tag{A.4}$$

where:

$$\lambda_{ij}(H) \equiv \sigma_{jj}\sum_{h=0}^{H-1}\left(e_i'B_h\Sigma e_j\right)^2 \quad \text{and} \quad \mu_i(H) \equiv \sum_{h=0}^{H-1}\left(e_i'B_h\Sigma B_h'e_i\right)$$

The denominator $\mu_i(H)$ corresponds to the total H-step-ahead forecast error variance for series i and the numerator is the forecast error variance for series i due to shocks in series j (normalised such that the shock is one standard deviation in size).

We compute each element $\psi_{kl}(H)$ in the transformed FEVD array as:

$$\psi_{kl}(H) = \frac{\sum_{i\in\mathscr{I}_k, j\in\mathscr{J}_l}\lambda_{ij}(H)}{\sum_{i\in\mathscr{I}_k}\mu_i(H)} \quad k,l = 1,\ldots K, H = 1,2,\ldots \tag{A.5}$$

where $\mathscr{I}_k$ and $\mathscr{J}_l$ are sets containing the row and column indexes respectively for the elements in the MF-VAR FEVD array that correspond to the element $\phi_{kl}(H)$ in the sense discussed above.

For the previous bivariate example, the elements in the MF-VAR FEVD array that correspond to $\phi_{kl}(H)$ are those contained in the sub-array $\Theta_{kl}(H)$ in Eq. (A.3). For example, for $k = 1, l = 1$, we have $\mathscr{I}_1 = \{1, \ldots, 4\}, \mathscr{J}_1 = \{1, \ldots, 4\}$ (the elements of $\Theta_{11}(H)$) and thus:

$$\psi_{11}(H) = \frac{\sum_{i\in\mathscr{I}_1, j\in\mathscr{J}_1} \lambda_{ij}(H)}{\sum_{i\in\mathscr{I}_1} \mu_i(H)} = \frac{\sum_{i=1,j=1}^{4} \lambda_{ij}(H)}{\sum_{i=1}^{4} \mu_i(H)} \qquad H = 1, 2, \ldots$$

Likewise, for $k = 2, l = 1$ we find $\mathscr{I}_2 = \{5\}, \mathscr{J}_1 = \{1, \ldots, 4\}$ (the elements of $\Theta_{21}(H)$) and thus:

$$\psi_{21}(H) = \frac{\sum_{i\in\mathscr{I}_2, j\in\mathscr{J}_1} \lambda_{ij}(H)}{\sum_{i\in\mathscr{I}_2} \mu_i(H)} = \frac{\sum_{j=1}^{4} \lambda_{5j}}{\mu_5(H)} \qquad H = 1, 2, \ldots$$

Appendix B. Data

B.1. Construction of weekly financial series

The MF-VAR approach of Ghysels (2016) is applicable in situations where the number of high-frequency time periods per low-frequency period varies deterministically over time. This does, however, somewhat complicate the implementation of the method. As discussed in the main text, we employ a combination of monthly macroeconomic and weekly financial time series for the empirical analysis. Working with monthly and weekly time intervals in a traditional sense, the number of weeks per calendar month varies from month to month.

To avoid the complications of deterministic time variation in the number of weeks per calendar month, we pre-process the data and work with what we term 'pseudo weeks' rather than standard calendar weeks. This approach is possible because we observe all financial series at a daily frequency that is higher than the final desired weekly frequency. These pseudo weeks are constructed by dividing the trading days within each month into four sub-periods whose lengths vary but are as close as possible to being equal. For example, months with 20 trading days are divided into four 5-day sub-periods, those with 19 trading days are divided into three 5-day periods and a 4-day period, those with 22 days are split into two 5-day periods and two 6-day periods and so on. The vast majority of pseudo-weeks contain either 5 or 6 trading days. However, Februarys or months with an unusually large number of weekday non-trading days due to holidays may have one or more weeks with four trading days.

With the exception of months containing exactly 20 trading days (always split into four weeks of equal length), how the weeks are ordered within a given month will influence the values (e.g. returns) obtained. For example, months with 21 working days can be split into four pseudo-weeks with lengths 5–5-5–6, with lengths 5–5-6–5, with lengths 5–6-5–5 or with lengths 6-5–5-5, each of which will produce different final values for prices, returns and return volatilities. Therefore to avoid this issue, we compute values over all possible split orders for a given month and then average the resulting values. Finally, while calculating pseudo-weekly returns or return volatilities for each period, we also adjust the return and volatility values obtained to account that the length of the return period differs slightly from one pseudo-week to another (4, 5 or 6 trading days).

B.2. Plots of data series

The US financial and macroeconomic series used to estimate spillover measures are plotted for the full sample period in Fig. A.1. Major economic and financial events during the sample period are clearly visible in the plots, either as substantial increases in financial volatility or large changes in the CFNAI; examples include the 1980–1981 recession, the Asian and Russian financial crisis and the collapse of Long-Term Capital Management in the late 1990s, the dotcom bubble and 9/11 in the early 2000s, the recent global financial crisis in the late 2000s and early 2010s, and the onset of the COVID-19 pandemic.

Appendix C. Supplementary empirical results

C.1. Additional plots of pairwise spillover measures

in Fig. A.2. Panel (a) plots financial to financial pairwise spillover estimates, panel (b) plots pairwise spillovers from each of the two financial markets to the real economy and panel (c) shows the reverse from the real economy to financial markets. Fig. A.2 complements the spillover decomposition plots of Fig. 2, with the latter being more suited to analysing the relative contribution of the individual spillover measures to the total level, and the former more suited to analysing the dynamics of specific spillover measures in absolute terms. In particular, Fig. A.2 allows us to more effectively examine how changes in pairwise spillovers relate to key historical events as we did previously for the total spillover index in Fig. 1.

There are several instances of increases in total spillovers being driven by increased spillovers from the financial sector to the real side of the economy. During the 1980 to 1982 inflationary period, spillovers from the bond markets to the real econ-

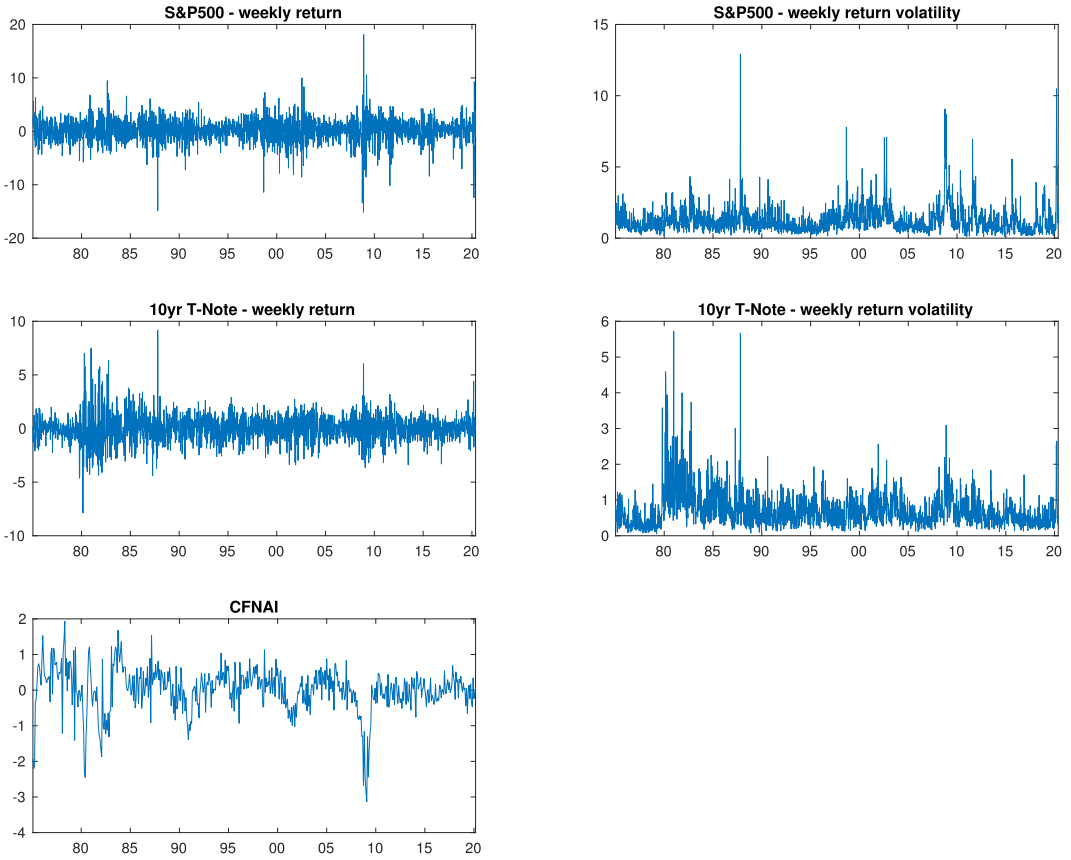

Fig. A.1. Time series plots of financial and real economy series. The financial (S&P500 and 10-year Treasury Bond) and real economy (CFNAI) series are plotted for the full sample period 1975:01 to 2020:06. Returns and return volatilities are expressed in percentage terms for the weekly frequency, with standard deviations plotted for the latter constructed using a range-based approach detailed in Appendix C.2. The CFNAI series is plotted in level form..

omy increased substantially. They reached local peaks in three instances: March 1980, December 1980 and February 1982, coinciding with federal funds target rate increases. These tightenings of monetary policy by the Volcker Fed led to higher interest rates, generating immediate contractionary impacts on the real economy, as captured by the spillovers from bond markets to the real economy reaching 8%. They subsequently declined gradually as inflation was brought under control leading to lower interest rates. These events are also visible from the mixed-frequency spillover measures in panel (a) of Fig. 2, but not the equivalent common-frequency measures in panel (b).

The Federal Reserve implemented another rate hike cycle between December 1986 and September 1987, with the federal funds target rate rising from 5.75% to 7.25% over this period. Consequently, spillovers from bond markets to the real economy increased from 3.2% in November 1986 to as high as 6.5% in May 1987. When the Federal Reserve lowered its policy rate back to 6.75% in response to Black Monday, spillovers from bonds to the real economy declined to 4% as of December 1987. However, the Federal Reserve returned to its rate hike cycle in March 1988 and increased the fed funds target rate to 9.75% by the end of 1988 and kept it there for another year. As a result, spillovers from the bond market to the real economy stayed above four percentage points until October 1990.

Increases in spillovers from equity markets to the real economy are visible after the Black Monday stock market crash in October 1987. The same occurred in the mid and late 1990s due to the build-up to the Mexican Tequila crisis beginning at the end of 1994, followed by the Asian and Russian financial crises and the collapse of LTCM in 1997 and 1998. During this period, spillovers from the stock and bond markets to the real economy increased from two to 6 percentage points. Both measures also increased in May 2000, when the Federal Reserve raised its policy rate from 6% to 6.5% despite the dot-com bubble bursting in the first half of 2000.

Following this, spillovers from equities and bonds to the real economy declined. However, from mid-2006, both recorded a sharp increase from around 2 to 8 percentage points when the Fed caught the markets off-guard by increasing the federal funds target rate in both May and June. With the start of the 2008–2009 financial crisis, spillovers from both bond and equity markets to the real economy increased sharply, with the latter being the most significant contributor to the sudden increase in the total spillover index throughout this crisis. The effects of the 2013 Taper Tantrum episode are also visible in panel (a)

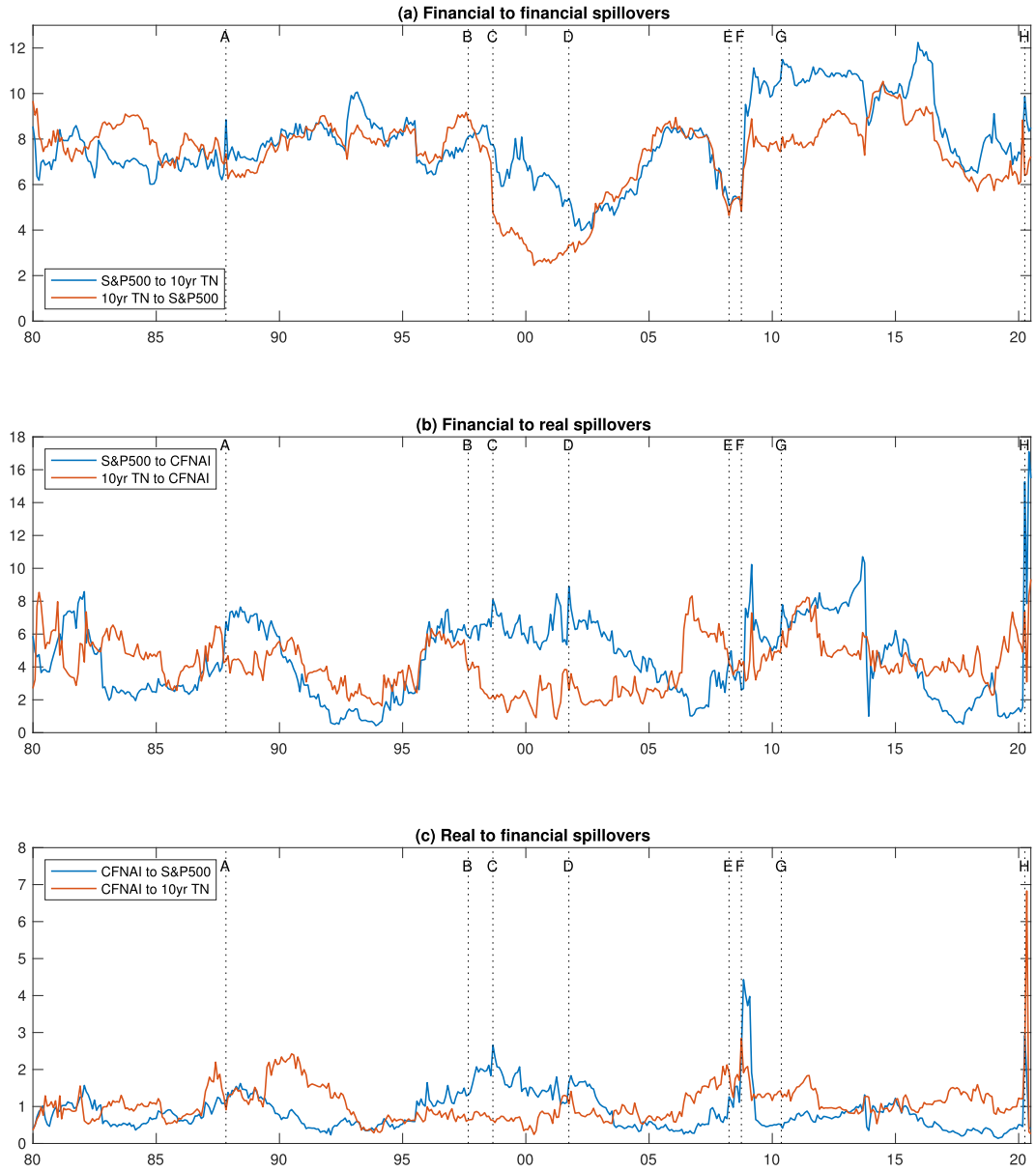

Fig. A.2. Pairwise macro-financial spillover measures. The figure presents plots of the pairwise macro-financial spillover measures for the sample period 1980:01 to 2020:06. All measures plotted are for the mixed-frequency case. Return levels are employed for the financial S&P500 and 10-year Treasury Note series, and levels for the real economy CFNAI series. Values are computed using a 3-month forecast horizon and a 60-month rolling window. The points marked are as follows. A: Black Monday, Oct '87, B: Asian financial crisis, Jul '97, C: Russian financial crisis and LTCM collapse, Aug to Sept '98, D: September 11, Sept '01, E: the collapse of Bear Stearns, Mar '08, F: Lehman Brothers collapse, AIG bailout and Fannie Mae and Freddie Mac being placed in government conservatorship, May '09, G: start of the EU debt crisis in April '10 and flash crash of May '10, and H: COVID-19 pandemic, March '20..

as a sharp increase in spillovers from the bond market to equities, reflecting the negative shock to bond market investor sentiment and its potential for spillovers to the equity market.

As discussed above, both the mixed-frequency and common-frequency total spillover indices in Fig. 1 responded similarly to the COVID-19 shock, with a significant spike in total spillovers. From Fig. A.2 it is clear that all of the individual components of spillovers increase in response to the pandemic. The largest increases in an absolute sense are noted for spillovers from financial markets to the real economy, particularly from the equity market. However, whilst smaller in absolute terms (reaching a level of just under 7%), in panel (c), we observe very large increases in spillovers from the real economy to bond markets relative to the pre-pandemic levels.

C.2. Total spillover indexes for logarithmic return volatilities

We also computed our mixed-frequency macro-financial spillover index using logarithmic return volatilities in place of return levels for the financial series as in Diebold and Yılmaz (2014). These previous studies employ the range-based esti-

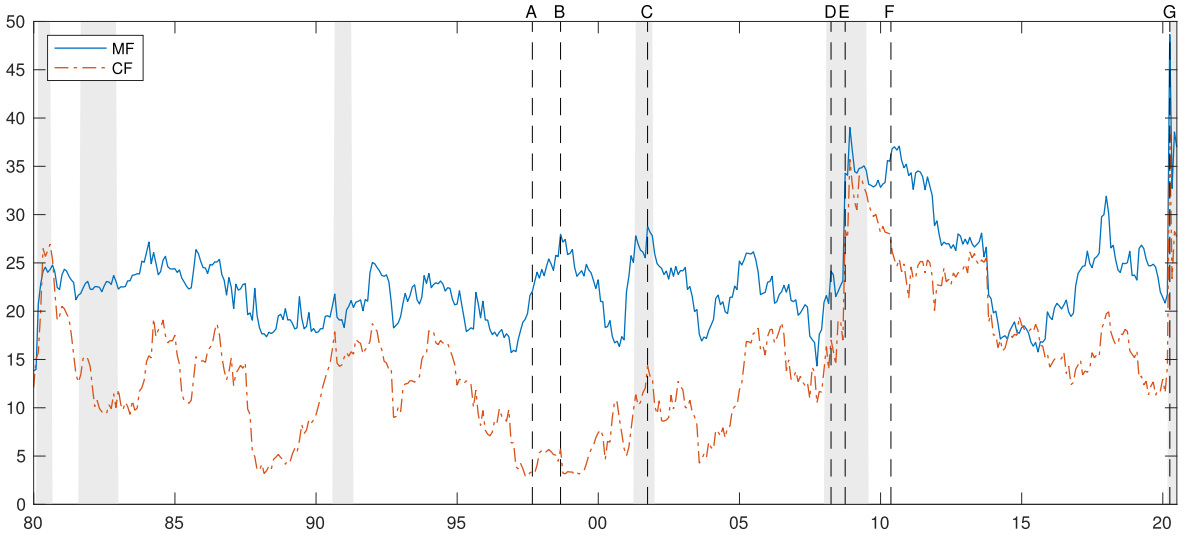

Fig. A.3. Total spillover indexes between financial and real economy series obtained from logarithmic return volatilities. Total spillover indexes for mixed-frequency (denoted MF) and common-frequency (denoted CF) approaches are presented for the sample period 1980:01 to 2020:06. Logarithmic return volatilities are employed for the financial S&P500 and 10-year Treasury Note series, levels for the real economy CFNAI series. Values are computed using a 3-month forecast horizon and a 60-month rolling window. The points marked are as follows. A: Asian financial crisis, Jul '97, B: Russian financial crisis and LTCM collapse, Aug to Sept '98, C: September 11, Sept '01, D: the collapse of Bear Stearns, Mar '08, E: Lehman Brothers collapse, AIG bailout and Fannie Mae and Freddie Mac being placed in government conservatorship, May '09, F: start of the EU debt crisis in April '10, flash crash of May '10 and G: COVID-19 pandemic, March '20. Shaded areas correspond to NBER US recession dates..

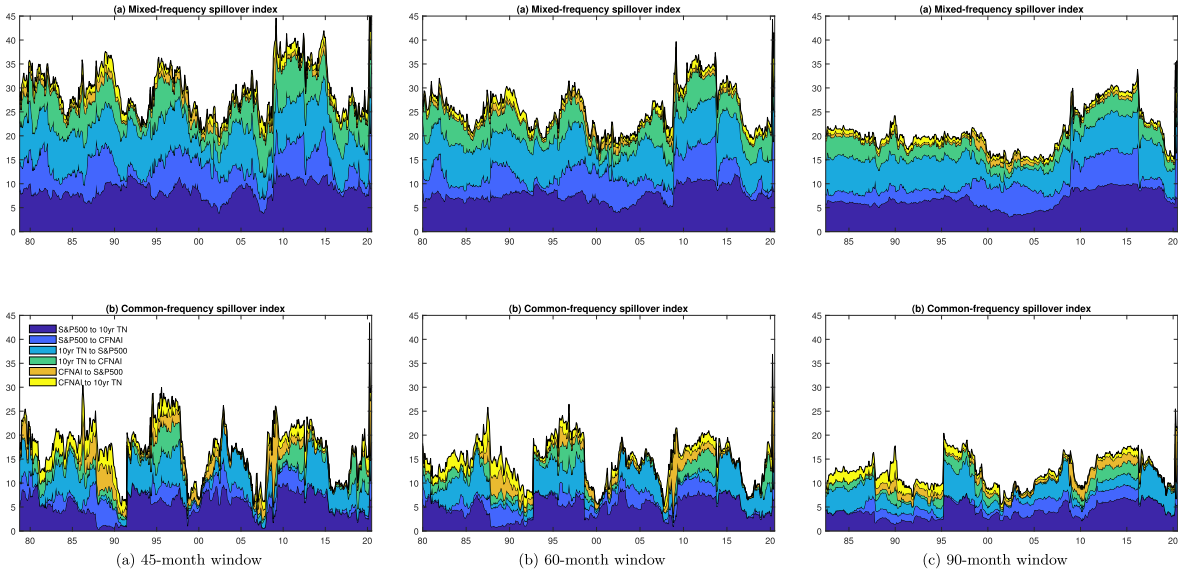

Fig. A.4. Impact of rolling estimation window length on spillover measures. The figure presents area plots in which the top of the complete shaded area corresponds to the relevant total spillover index and each shaded area below representing the contribution of each specific pairwise spillover to the value of the total spillover index. The relevant spillover measures for the mixed-frequency case are plotted in the top row and those for the common-frequency case in the bottom row. Return levels are used for the financial S&P500 and 10-year Treasury Note series and levels for the real economy CFNAI series. Values are computed using a 3-month forecast horizon and 45-month (in column 1), 60-month (in column 2) or 90-month (in column 3) rolling windows for the sample period 1980:01 to 2020:06..

mator of return volatility proposed by Parkinson (1980), which estimates return volatility from the high and low prices during the chosen return period. As daily data on high and low prices are not available during the earlier parts of the sample period, we approximate this estimator by replacing the high, and low prices with the highest and lowest daily closing prices observed each week. For the later parts of the sample period where both estimators can be computed, we confirm that our volatility measure using only close prices is highly correlated with the Parkinson (1980) estimator, with a correlation coefficient of just over 0.9 for the S&P500 and just under 0.9 for the 10-year Treasury Note series.

Fig. A.3 presents an equivalent plot to Fig. 1, for the case where logarithmic return volatilities are employed instead of return levels for financial series. The total spillover index series display broadly similar dynamics to those observed for the case of return levels. For the case of return volatilities, the spillover indexes typically display slightly more pronounced spikes around adverse financial or economic events than the ones for return levels (particularly during the Asian and Russian financial crises). Financial volatility spillovers typically rise in turbulent times and drop in tranquil times, whereas return spillovers may increase in both situations. Finally, we again observe similar differences in the level of macro-financial spillovers between the mixed-frequency and common-frequency cases, with the former implying a higher average level of spillovers over the sample period.

C.3. Impact of rolling estimation window size

Fig. A.4 illustrates how the choice of rolling window length employed to estimate the spillover measures affects the spillover measures obtained. It can be seen that the key results of Section 3.2 are qualitatively unchanged, with mixed-frequency spillovers consistently higher than common frequency equivalents, spillovers of financial origin making up the majority of total spillovers, and the indexes increasing during periods of economic or financial turbulence. As expected for any type of rolling window estimation approach, longer (shorter) window lengths result in smoother (rougher) estimated indexes.

References

Acharya, V.V., Pedersen, L.H., Philippon, T., Richardson, M., 2016. Measuring systemic risk. Rev. Finan. Stud. 30, 2–47.
Adams, Z., Füss, R., Gropp, R., 2014. Spillover effects among financial institutions: a state-dependent sensitivity value-at-risk approach. J. Finan. Quant. Anal. 49, 575–598.
Adrian, T., Brunnermeier, M.K., 2016. CoVaR. Am. Econ. Rev. 106, 1705–1741.
Allen, L., Bali, T.G., Tang, Y., 2012. Does systemic risk in the financial sector predict future economic downturns? Rev. Finan. Stud. 25, 3000–3036.
Andersen, T., Bollerslev, T., Diebold, F., Vega, C., 2003. Micro effects of macro announcements: Real-time price discovery in foreign exchange, typescript. Am. Econ. Rev. 93, 38–62.
Ando, T., Greenwood-Nimmo, M., Shin, Y., 2022. Quantile connectedness: modeling tail behavior in the topology of financial networks. Manage. Sci. 68, 2377–3174.
Barunik, J., Kočenda, E., Vácha, L., 2015. Asymmetric connectedness on the U.S. stock market: Bad and good volatility spillovers. J. Finan. Mark., 1–49
Barunik, J., Křehlík, T., 2018. Measuring the frequency dynamics of financial connectedness and systemic risk. J. Finan. Econ. 16, 271–296.
Baur, D.G., 2012. Financial contagion and the real economy. J. Bank. Finance 36, 2680–2692.
Bernanke, B.S., Gertler, M., Gilchrist, S., 1996. The financial accelerator and the flight to quality. Rev. Econ. Stat. 78, 1–15.
Billio, M., Getmansky, M., Lo, A.W., Pelizzon, L., 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. J. Financ. Econ. 104, 535–559.
Brenner, M., Pasquariello, P., Subrahmanyam, M., 2009. On the Volatility and Comovement of U.S. financial markets around macroeconomic news announcements. J. Finan. Quant. Anal., 44:1265
Brownlees, C.T., Engle, R.F., 2017. SRISK: A conditional capital shortfall measure of systemic risk. Rev. Finan. Stud. 30, 48–79.
Brunnermeier, M.K., Gorton, G., Krishnamurthy, A., 2011. Risk topography. NBER Macroecon. Annual 26, 149–176.
Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Rev. Finan. Stud. 21, 1509–1531.
Cingano, F., Manaresi, F., Sette, E., 2016. Does credit crunch investment down? New evidence on the real effects of the bank-lending channel. Rev. Finan. Stud. 29, 2737–2773.
Claessens, S., Tong, H., Wei, S.-J., 2012. From the financial crisis to the real economy: Using firm-level data to identify transmission channels. J. Int. Econ. 88, 375–387.
Claeys, P., Vašíček, B., 2014. Measuring bilateral spillover and testing contagion on sovereign bond markets in Europe. J. Bank. Finance 46, 151–165.
Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. J. Econ. 138, 291–311.
Demirer, M., Diebold, F.X., Liu, L., Yılmaz, K., 2018. Estimating global bank network connectedness. J. Appl. Econ. 33, 1–15.
Diebold, F.X., 2015. Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of Diebold–Mariano Tests. J. Bus. Econ. Stat. 33, 1–9.
Diebold, F.X., Yılmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. Econ. J. 119, 158–171.
Diebold, F.X., Yılmaz, K., 2012. Better to give than to receive: Predictive directional measurement of volatility spillovers. Int. J. Forecast. 28, 57–66.
Diebold, F.X., Yılmaz, K., 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. J. Econ. 182, 119–134.
Ellington, M., 2018. Financial market illiquidity shocks and macroeconomic dynamics: Evidence from the UK. J. Bank. Finance 89, 225–236.
Engle, R.F., Jondeau, E., Rockinger, M., 2015. Systemic Risk in Europe. Rev. Finance 19, 145–190.
Galvão, A.B., Owyang, M.T., 2018. Financial stress regimes and the macroeconomy. J. Money, Credit Bank. 50, 1479–1505.
Ghysels, E., 2016. Macroeconomics and the reality of mixed frequency data. J. Econ. 193, 294–314.
Giglio, S.W., Kelly, B., Pruitt, S., 2016. Systemic risk and the macroeconomy: An empirical evaluation. J. Financ. Econ. 119, 457–471.
Green, T.C., 2004. Economic news and the impact of trading on bond prices. The Journal of Finance 59, 1201–1233.
Hubrich, K., Tetlow, R.J., 2015. Financial stress and economic dynamics: The transmission of crises. Journal of Monetary Economics 70, 100–115.
Ivashina, V., Scharfstein, D., 2010. Bank lending during the financial crisis of 2008. J. Financ. Econ. 97, 319–338.
Lenza, M., and G.E. Primiceri. 2020. How to Estimate a VAR after March 2020. Working Paper.
Li, D., Magud, N.E., Valencia, F., 2019. Financial Shocks and Corporate Investment in Emerging Markets. Journal of Money, Credit and Banking 52, 613–644.
Parkinson, M., 1980. The Extreme Value Method for Estimating the Variance of the Rate of Return. The Journal of Business 53, 61–65.
Paye, B.S., 2012. 'Déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. J. Financ. Econ. 106, 527–546.
Pesaran, M.H., Shin, Y., 1998. Generalized impulse response analysis in linear multivariate models. Economics Letters 58, 17–29.
Prieto, E., Eickmeier, S., Marcellino, M., 2016. Time Variation in Macro-Financial Linkages. Journal of Applied Econometrics 31, 1215–1233.

Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Rev. Finan. Stud. 23, 821–862.
Savor, P., Wilson, M., 2013. How much do investors care about macroeconomic risk? Evidence from scheduled economic announcements. J. Finan. Quant. Anal. 48, 343–375.
Schorfheide, F., Song, D., 2015. Real-time forecasting with a mixed-frequency VAR. J. Bus. Econ. Stat. 33, 366–380.
Schorfheide, F., Song, D., 2020. Real-time forecasting with a (standard) mixed-frequency VAR During a Pandemic. Working Paper.