International Journal of Forecasting 40 (2024) 777–795

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Financial-cycle ratios and medium-term predictions of GDP: Evidence from the United States

Graziano Moramarco[1]

University of Bologna, Department of Economics, Piazza Scaravilli 2, 40126 Bologna, Italy

ARTICLE INFO

Keywords:
Financial cycle
Housing market
Firm debt
GDP forecasts
Crisis
Price–rent ratio
Medium term

ABSTRACT

Using a large quarterly macroeconomic dataset for the period 1960–2017, we document the ability of specific financial ratios from the housing market and firms' aggregate balance sheets to predict GDP over medium-term horizons in the United States. A cyclically adjusted house price-to-rent ratio and the liabilities-to-income ratio of the non-financial non-corporate business sector provide the best in-sample and out-of-sample predictions of GDP growth over horizons of one to five years, based on a wide variety of rankings. Small forecasting models that include these indicators outperform popular high-dimensional models and forecast combinations. The predictive power of the two ratios appears strong during both recessions and expansions, stable over time, and consistent with well-established macro-finance theory.

1. Introduction

Following the Global Financial Crisis and the Great Recession of 2008–2009, the interactions between asset prices, households' and firms' balance sheets, and real economic activity have become increasingly important in macroeconomic analysis (see, e.g., Gertler and Gilchrist 2018, Mian and Sufi 2018, among many others). These interactions give rise to boom-and-bust financial cycles (Borio 2014, Claessens et al. 2011, Drehmann et al. 2012), with major repercussions on business cycles. In particular, house prices and credit have received a great deal of attention and have been argued to provide "the most parsimonious description of the financial cycle" (Borio, 2014). Empirical research has delivered substantial results on the predictive potential of financial-cycle indicators for business cycles (e.g., Adrian et al. 2019, Jordà et al. 2016, Mian et al. 2017). However, comprehensive comparative evaluations are still needed to determine whether any specific variables, among those related to credit, balance sheets, and housing, stand out as effective predictors of economic activity.

This paper offers novel evidence on the ability of specific financial-cycle indicators to predict GDP over the medium term in the United States. Based on an extensive evaluation using a dataset of 262 quarterly macroeconomic and financial variables for the period 1960–2017, we find that two ratios provide the best in-sample and out-of-sample predictions of GDP growth over horizons of one to five years: a cyclically adjusted house price-to-rent (CAPR) ratio, calculated over the aggregate stock of owner-occupied housing (Contessi and Kerdnunvong 2015, Davis et al. 2008); and the non-financial non-corporate business sector liabilities-to-income (NNBLI) ratio. The CAPR ratio is a robust valuation metric for the housing market, representing the housing counterpart of the popular cyclically adjusted price-to-earnings (CAPE) ratio introduced by Campbell and Shiller (1998) for the stock market, while the NNBLI ratio measures the debt burden of non-corporate (small) businesses, which represent the vast majority (more than 80% in recent years) of

E-mail address: graziano.moramarco@unibo.it.

[1] I wish to thank the editor, an associate editor, and two anonymous reviewers for useful comments and suggestions. The usual disclaimer applies. I gratefully acknowledge financial support from the Italian Ministry of Education, Universities and Research (MIUR), PRIN 2017 Grant 2017TA7TYC.

https://doi.org/10.1016/j.ijforecast.2023.05.007

firms in the United States and account for almost 20% of total revenues and value added.[2] To our knowledge, this is the first paper that shows the outstanding predictive power of these ratios.

Both the CAPR and the NNBLI ratios are inversely related with subsequent medium-term economic activity. CAPR appears especially effective at predicting GDP growth over three to five years, while NNBLI provides its best predictions over horizons of one to three years. Fig. 1 plots the cumulative growth rate of GDP relative to 20 quarters before and compares it with the (log) CAPR ratio lagged by 20 quarters, while Fig. 2 compares the cumulative growth rate of GDP over the last 12 quarters with the 12-quarter lag of the NNBLI ratio. Both figures show a strong negative correlation (-0.7 and -0.61, respectively).

We first provide a set of baseline results, based on both in-sample and out-of-sample evaluations, using both direct forecasts produced by autoregressive distributed lag (ARDL) models and iterated forecasts by vector autoregressive (VAR) models. These results offer unambiguous evidence of the special importance of CAPR and NNBLI, compared to all other predictors, and formalize the intuition provided by Figs. 1 and 2. We also find that forecasts produced by one-predictor models using the best financial-cycle ratios outperform forecasts by high-dimensional models and forecast combinations. This appears remarkable, since a large body of literature has shown that small forecasting models are often outperformed by models and methods that exploit a large amount of information, such as large Bayesian VARs (see, e.g., Bańbura et al. 2010, Carriero et al. 2015, Koop 2013, Koop and Korobilis 2013), factor models (e.g., Forni et al. 2001, Stock and Watson 2002, 2011), and forecast combinations (e.g., Stock and Watson 2003, 2004).

Next, we present a variety of extensions and robustness checks, including the estimation of quantile regressions for GDP growth, several checks on forecast instabilities, forecasts by time-varying-parameter (TVP) models, a comparison of alternative variable-selection methods, and the evaluation of forecasts based on real-time (unrevised) data vintages, when available. This extensive analysis provides several other striking results, which can be summarized as follows: the predictive content of the two ratios is not limited to recessions or periods of financial distress but is much more general. First, quantile regressions reveal that their strong relationship with GDP growth over the medium term is stable across different parts of the GDP growth distribution, particularly in the case of CAPR. This distinguishes the two ratios from general indicators of financial conditions, such as the National Financial Conditions Index (NFCI), which tend to be good predictors of the left tail of GDP but are not very useful for the rest of the distribution, as shown by the recent literature on growth at risk (Adrian et al. 2019, 2022) and confirmed by our results. Also, while economic forecasts are generally found to be affected by substantial instabilities (e.g., Clements and Hendry 2006, Rossi 2021, Stock and Watson 2003), the predictive power of CAPR and NNBLI appears quite stable over time. In particular, the two ratios are shown to be top predictors before, during, and after the Global Financial Crisis.

The paper is related to several strands of the literature. First, it follows other comparative evaluations of predictors of economic activity based on large datasets (e.g., Banerjee et al. 2005, Marcellino et al. 2003, Stock and Watson 2003). A large number of papers have focused on the role of financial variables, finding mixed evidence on their ability to forecast GDP (see, e.g., Claessens and Kose 2017, Stock and Watson 2003). Some variables, such as the term spread of interest rates, exhibit good forecast performance but only in specific periods (Chauvet and Potter 2013, Stock and Watson 2003). As mentioned above, the growth-at-risk literature (Adrian et al. 2019, 2022) shows that aggregate financial conditions help forecast tail risks to GDP. More specifically, our paper contributes to the literature on housing and credit cycles, and their predictive relationships with the business cycle. Leamer (2007, 2015) claims that "housing *is* the business cycle", showing that it is the economic sector with the largest contribution to U.S. recessions and offers important early warnings. Housing variables, such as building permits and housing starts, have long been used as leading indicators of GDP (Coulson and Kim 2000, Green 1997). Also, there is extensive evidence on the prominent role of housing wealth shocks in the Great Recession and the subsequent slow recovery (e.g., Mian et al. 2013, Mian and Sufi 2014). Credit booms, especially those driven by mortgage credit, have been followed by deep recessions, slow recoveries, and financial turmoil in recent decades (Jordà et al. 2013, 2016, 2017, Mian and Sufi 2018). In particular, high credit-to-GDP and household debt-to-GDP ratios predict lower GDP growth in the medium run (Mian et al., 2017) and financial instability (Borio & Lowe, 2002). Non-financial leverage has been a good predictor of growth vulnerability during the Great Recession (Reichlin et al., 2020). Adrian et al. (2022) show that loose financial conditions increase downside risks to GDP, especially over medium-term horizons (one to three years) and when credit-to-GDP growth is rapid. Finally, our results on the strong predictive ability of the CAPR and NNBLI ratios appear to be consistent with theoretical frameworks in which credit market conditions and (housing-based) collateralized borrowing by firms and households are major drivers of economic fluctuations, including the popular financial accelerator model by Bernanke et al. (1999) and more recent macro-finance models that explicitly incorporate the housing sector (e.g., Favilukis et al. 2017). At the end of the paper, we discuss in more detail the connections between our empirical findings and the theoretical insights provided by macro-finance research.

The remainder of the paper is organized as follows: Section 2 introduces the data, Section 3 presents the baseline results of in-sample and (pseudo-) out-of-sample evaluations, Section 4 presents the extensions and robustness checks, Section 5 discusses the macro-finance models that help explain the predictive power of CAPR and NNBLI, and Section 6 concludes.

[2] Sources: U.S. Internal Revenue Service, https://www.irs.gov/statistics/soi-tax-stats-integrated-business-data; U.S. Bureau of Economic Analysis, https://www.bea.gov/data/gdp/gross-domestic-product.

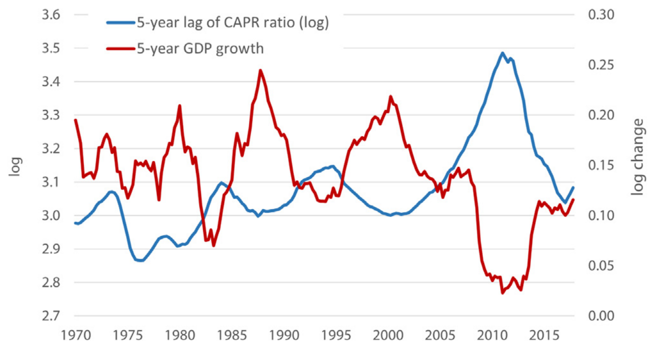

Fig. 1. CAPR ratio and five-year GDP growth.
Notes: The figure shows the 20-quarter lag of the log cyclically adjusted house price-to-rent (CAPR) ratio (left axis) and the cumulative growth rate of GDP over the previous 20 quarters (right axis) in the United States, from 1970Q1 to 2017Q4. House prices and rents are measured by Davis et al. (2008), and the data are available at https://www.aei.org/historical-land-price-indicators/.

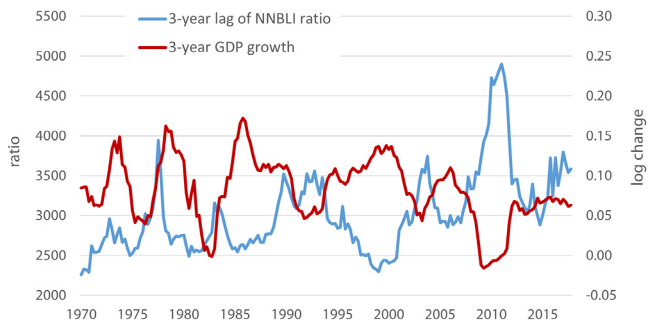

Fig. 2. NNBLI ratio and three-year GDP growth.
Notes: The figure shows the 12-quarter lag of the ratio of non-financial non-corporate business sector liabilities to disposable business income (NNBLI; divided by 100, left axis) and the cumulative growth rate of GDP over the previous 12 quarters (right axis) in the United States, from 1970Q1 to 2017Q4. The data are from the FRED-QD dataset (McCracken & Ng, 2021).

2. Variables and data

We conduct our analysis using a large quarterly dataset (262 macroeconomic and financial variables) over the period 1960Q1–2017Q4. The dataset combines the FRED-QD dataset[3] with the variables reported in Table 1, which include a variety of financial-cycle indicators (total credit to the non-financial sector, the credit–GDP ratio, household mortgage debt and the mortgage–income ratio, real house price growth and price–rent ratios, the ratio of residential fixed investment to GDP, real stock market prices and the cyclically adjusted price–earnings ratio, and household interest payments as a fraction of disposable income), as well as leading indicators (the OECD composite leading indicator and business confidence index) and the comprehensive National Financial Conditions Index (NFCI) by the Chicago Fed, capturing a variety of other financial factors. Table 1 also indicates the data transformations used for these additional variables. The FRED-QD dataset is made of 248 variables, which cover in detail a wide spectrum of macro areas. McCracken and Ng (2021) classify the variables into 14 groups: National Income and Product Accounts (NIPA); Industrial Production; Employment and Unemployment; Housing; Inventories, Orders, and Sales; Prices; Earnings and Productivity; Interest Rates; Money and Credit; Household Balance Sheets; Exchange Rates; Stock Markets; Non-Household Balance Sheets; and Other. For all variables in FRED-QD, we use the data transformations suggested by McCracken and Ng (2021), the only adaptation being the use of year-on-year instead of quarter-on-quarter changes/growth rates. Besides, for all rates and ratios (interest rates and spreads, unemployment rates, exchange rates, balance-sheet ratios, etc.) for which differencing is suggested, we also consider the levels, i.e., non-transformed data. Table 2 lists the subset of FRED-QD variables that will be reported among the best predictors of GDP in the baseline results. In all results, uppercase labels identify variables from FRED-QD and lowercase labels the additional set of variables, as shown in Tables 1 and 2.

In analogy with the popular cyclically adjusted price-to-earnings ratio (CAPE) proposed by Campbell and Shiller (1998) for the stock market, which is calculated by dividing the real S&P 500 index by the 10-year moving average of real earnings on the index, the CAPR ratio is calculated by dividing real house prices by the 10-year average real

[3] Available at https://research.stlouisfed.org/econ/mccracken/fred-databases The FRED-QD dataset was officially launched in May 2018. This paper uses the 2018-06 vintage and the associated labels for variables. In Section 4.8, we present results using historical (real-time) data vintages for a subset of variables for which they are available.

Table 1
Variables in addition to the FRED-QD dataset.

Label	Description	Data source	Source label	Transformation
bci	OECD Business Confidence Index	OECD	–	x_t
cli	OECD Composite Leading Indicator	OECD	–	x_t
sp500	Real S&P500 index growth	Shiller	–	$\ln(x_t) - \ln(x_{t-4})$
cape	Cyclically adjusted price/earnings ratio	Shiller	–	$\ln(x_t)$
cred	Real credit growth	FRED	CRDQUSAPABIS/CPIAUCSL	$\ln(x_t) - \ln(x_{t-4})$
cred_gdp	Credit/GDP ratio	BIS	–	x_t
hpi	Real house price growth	Shiller	–	$\ln(x_t) - \ln(x_{t-4})$
pr	Price/rent ratio	AEI	–	$\ln(x_t)$
capr	Cyclically adjusted price/rent ratio	AEI; FRED	– ; CUUR0000SEHA, CPIAUCNS	$\ln(x_t)$
mortg	Household real mortgage debt growth	FRED	HHMSDODNS/CPIAUCSL	$\ln(x_t) - \ln(x_{t-4})$
mortg_inc	Household mortgage/income ratio	FRED	HHMSDODNS/DSPI	x_t
prfi_gdp	Private residential fixed investment/GDP ratio	FRED	PRFI/GDP	x_t
pip_inc	Personal interest payments/income ratio	FRED	B069RC1/DSPI	x_t
nfci	Chicago Fed national financial condition index	FRED	NCFI	x_t

Notes: Shiller = http://www.econ.yale.edu/~shiller/data.htm; AEI = American Enterprise Institute, https://www.aei.org/historical-land-price-indicators/, based on original data by Davis et al. (2008) (using the Case–Shiller price index after 2000).

Table 2
Best-performing FRED-QD variables.

Label	Description
AAA	Moody's Seasoned Aaa Corporate Bond Yield© (Percent)
AHETPIx	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Total
AMDMUOx	Real Value of Manufacturers' Unfilled Orders for Durable Goods Industries (Millions of 2012 Dollars), deflated by Core PCE
B021RE1Q156NBEA	Shares of gross domestic product: Imports of goods and services (Percent)
CONSPIx	Nonrevolving consumer credit to Personal Income
DOTSRG3Q086SBEA	Personal consumption expenditures: Other services (chain-type price index)
FEDFUNDS	Effective Federal Funds Rate (Percent)
GS1TB3Mx	1-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market
HNOREMQ027Sx	Real Real Estate Assets of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE
IPCONGD	Industrial Production: Consumer Goods (Index 2012=100)
IPNCONGD	Industrial Production: Nondurable Consumer Goods (Index 2012=100)
ISRATIOx	Total Business: Inventories to Sales Ratio
LIABPIx	Liabilities of Households and Nonprofit Organizations Relative to Personal Disposable Income (Percent)
MORTGAGE30US	30-Year Conventional Mortgage Rate© (Percent)
NNBTILQ027SBDIx	Non-financial Non-corporate Business Sector Liabilities to Disposable Business Income (Percent)
NWPIx	Net Worth of Households and Nonprofit Organizations Relative to Disposable Personal Income (Percent)
PCECC96	Real Personal Consumption Expenditures (Billions of Chained 2009 Dollars)
PRFIx	Real private fixed investment: Residential (Billions of Chained 2009 Dollars), deflated using PCE
REVOLSLx	Total Real Revolving Credit Owned and Securitized, Outstanding (Billions of 2012 Dollars), deflated by Core PCE
S&P: div yield	S&P's Composite Common Stock: Dividend Yield
TLBSNNCBBDIx	Non-financial Corporate Business Sector Liabilities to Disposable Business Income (Percent)
UNRATESTx	Unemployment Rate less than 27 weeks (Percent)
USMINE	All Employees: Mining and logging (Thousands of Persons)

Notes: The table reports the subset of variables from the FRED-QD dataset (McCracken & Ng, 2021) that rank among the best predictors of GDP in this paper (see Sections 3.1 and 3.2). See Section 2 for more details on data transformations. Starting in the FRED-QD vintage 2018-09, the label NNBTILQ027SBDIx has been replaced with TLBSNNBBDIx.

rents, i.e.:

$$CAPR_t = \frac{HPI_t}{\frac{1}{40}\sum_{i=1}^{40} R_{t-1}}$$

where HPI_t denotes the house price index in real terms at time t (in quarters), and R_t is the imputed real rent at time t. Following Contessi and Kerdnunvong (2015), we calculate the CAPR using data by Davis et al. (2008), who measured house prices and (imputed) rents on the entire stock of owner-occupied houses, representing the large majority of houses in the United States.[4] Contessi and Kerdnunvong (2015) use the ratio to test for bubbles in the housing market.

The NNBLI ratio is provided by the FRED-QD dataset and is calculated as the ratio of liabilities to business disposable income. The non-corporate business sector primarily consists of partnerships and sole proprietorships, which represent the typical organizational forms of small firms.[5]

[4] See Davis et al. (2008) for methodological details. The raw data on house prices and rents over the period 1960Q1–2018Q2 are available at https://www.aei.org/historical-land-price-indicators/. To calculate the 40-quarter moving average of rents from 1960Q1 to 1970Q1, we extrapolate the rent series backward from 1960Q1 to 1950Q1 using the growth rate of the rent of primary residence, measured by the U.S. Bureau of Labor Statistics (BLS) and provided by FRED at https://fred.stlouisfed.org/. The home-ownership rate in the United States in the period from 1965 to 2017 ranged between 63% and 69% (see https://www.census.gov/housing/hvs/index.html).

[5] Non-corporate businesses are pass-through firms. That is, their income is passed on to firm owners and treated (taxed) as personal

3. Baseline results

3.1. In-sample evaluation

Following Stock and Watson (2003), we conduct our in-sample evaluation of predictive power using autoregressive distributed lag (ARDL) models for multi-period cumulative GDP growth rates. More specifically, we predict the h-quarter-ahead cumulative GDP growth rate, with $h = 4, 12, 20$, using a bivariate model that includes one predictor at a time plus lags of GDP growth. We also considered results for $h = 28$, which corroborate those obtained for $h = 20$ and are not reported in the interest of space. The economic significance of alternative predictors is then evaluated using the R^2 of the regressions.

The ARDL regressions are as follows:

$$y_t^h = \beta_0 + \beta_1(L)y_{t-h} + \beta_{2,i}(L)x_{i,t-h} + u_t \tag{1}$$

where y_t^h is the log approximation of the cumulative GDP growth rate over a period of length h, i.e., $y_t^h = \ln(GDP_t) - \ln(GDP_{t-h})$; y_t is the log approximation of the year-on-year GDP growth rate at time t, i.e., $y_t = \ln(GDP_t) - \ln(GDP_{t-4})$; $x_{i,t}$ is the ith candidate predictor; u_t is the error term; β_0 is a constant; and $\beta_1(L)$ and $\beta_{2,i}(L)$ are lag polynomials, such that $\beta_1(L)y_{t-h} = \sum_{j=1}^{p} \beta_{1,j}y_{t-h+1-j}$ and $\beta_{2,i}(L)x_{i,t-h} = \sum_{j=1}^{q} \beta_{2,i,j}x_{i,t-h+1-j}$.

The models are estimated using data in the time range from 1974Q1 to 2017Q4. The starting date is chosen so as to ensure that 95% of the time series have no missing data in the sample (shorter series are excluded here but will be used in the out-of-sample evaluation of Section 3.2). To ensure perfect comparability of R^2, we fix the lag length across models. In particular, both p and q are set at 5 to adequately account for serial correlation in quarterly time series.

Table 3 reports the R^2 of the regressions. The predictors are listed in descending order of R^2 and only the best 10 are reported for each value of h. The log cyclically adjusted price–rent ratio (label: *capr*) dominates over longer horizons. It is the best predictor for $h = 12, 20$ and the second-best for $h = 4$. The OECD composite leading indicator (*cli*) is a particularly strong predictor over short horizons. It ranks first for $h = 4$, and then it gradually falls down the ranking as the horizon increases (fifth for $h = 12$ and out of the top 10 for $h = 20$). Most of the top positions are occupied by financial-cycle indicators. The NNBLI ratio (FRED-QD label: *NNBTILQ027SBDIx*[6]) is the third-best predictor for $h = 4$ and for $h = 12$, after CAPR and the unadjusted price–rent ratio (*pr*), and the fifth-best predictor for $h = 20$. Private residential fixed investment, both as a share of GDP and in growth rates (labels: *prfi_gdp* and *PRFIx*, respectively), performs especially well over the four-quarter horizon, while the mortgage–income ratio (*mortg_inc*), household liabilities and net worth relative to disposable income (*LIABPIx* and *NWPIx*), and the credit–GDP ratio (*cred_gdp*) are effective predictors over longer horizons.

Among the other variables, the unfilled orders for nondurable goods (*AMDMUOx*) feature in the top 10 for $h = 4, 12$, while the inventories to sales ratio (*ISRATIOx*) for $h = 12, 20$.

The usefulness of the CAPR appears even more remarkable if the lag orders p and q are selected by the Bayes information criterion (BIC; we first select p, then q conditional on p, given the maximum lag length of five). In this case, it is the best predictor at all horizons (results are not reported).

Finally, the results on CAPR are not simply determined by the specific indices of house prices and rents considered. To check this, we replicate the analysis on two alternative measures of CAPR, one calculated using the house price index by Shiller (2015) and the average rent of primary residences in U.S. cities by the Bureau of Labor Statistics, the second using the nominal house price and rent indices by the OECD. Taking these two measures, CAPR still ranks among the top predictors for all horizons. However, the best results are obtained using the data on the aggregate stock of owner-occupied housing by Davis et al. (2008).

3.2. Out-of-sample evaluation

The second part of the analysis evaluates the forecasting power of the predictors out of the estimation sample. To this aim, we track direct and iterated forecasts over time using a recursive-window scheme.[7] Direct forecasts are made using multi-period ARDL models, while iterated forecasts are produced by bivariate VAR models. The forecast horizons are the same as in the in-sample evaluation, i.e., 4, 12, and 20 quarters (again, $h = 28$ was considered but is not reported, as it only strengthens the main results for $h = 20$). Time series of forecasts over different horizons are constructed for each one of the competing models and then used for comparison. In particular, predictors are evaluated using the mean square forecast errors (MSFEs) computed over the period 1990Q1–2017Q4. Given the maximum forecast horizon of 20 quarters, this implies setting the ending point of the shortest estimation window to 1985Q1, which is accommodated by moving the starting point to an earlier date than in the in-sample evaluation, namely 1968Q2.[8]

income. As a result, by construction, the sector has zero disposable income in national accounts (see the Fed's Financial Accounts of the United States at https://www.federalreserve.gov/releases/z1/ and the Integrated Macroeconomic Accounts by the U.S. Bureau of Economic Analysis at https://www.bea.gov/data/special-topics/integrated-macroeconomic-accounts.) To calculate business disposable income at the denominator of the NNBLI ratio, FRED-QD uses data on corporate cash flows and taxes, which provide a good proxy.

6 Starting in the FRED-QD vintage 2018-09, this label is replaced by *TLBSNNBBDIx*.

7 As explained in Section 4.3, we also consider rolling windows of various fixed lengths. However, the best forecasts are generally achieved in the recursive-window case, so we use this for our baseline results.

8 As the maximum number of lags is set to five, this specific start date is chosen to ensure that all the VAR models are estimated using data as far back as 1967Q1, which is the first quarter in which data are available (after transformation) for at least 90% of the time series in the dataset. Concerning the multi-step ARDL models, the range of data used for estimation depends on the relevant horizon (in the case $h = 28$, estimation uses data as far back as 1960Q1). Also,

Table 3
Regression R^2 of single-predictor ARDL models.

	$h = 4$		$h = 12$		$h = 20$	
1	cli	0.43	capr	0.63	capr	0.68
2	capr	0.40	pr	0.57	pr	0.65
3	NNBTILQ027SBDIx	0.36	NNBTILQ027SBDIx	0.56	LIABPIx	0.58
4	prfi_gdp	0.35	AMDMUOx	0.55	mortg_inc	0.58
5	pr	0.35	cli	0.50	NNBTILQ027SBDIx	0.55
6	DOTSRG3Q086SBEA	0.33	mortg_inc	0.45	ISRATIOx	0.53
7	nfci	0.33	LIABPIx	0.44	cred_gdp	0.53
8	PRFIx	0.33	ISRATIOx	0.44	NWPIx	0.47
9	AMDMUOx	0.32	cred_gdp	0.42	AAA	0.47
10	IPCONGD	0.31	NWPIx	0.41	CONSPIx	0.44

Notes: For each h, the dependent variable is GDP growth over h quarters (cumulative). Models are estimated using data from 1974Q1 to 2017Q4. Please refer to Tables 1 and 2 for a description of the variables.

Competing models are initially estimated on the shortest sample 1968Q2–1985Q1 and used to make forecasts for the period from 1986Q1 (four-quarter-ahead forecast) to 1990Q1 (20-quarter-ahead forecast). Then the sample is recursively expanded by one quarter at a time, and the estimation and forecasting steps are repeated in each iteration (i.e., in the second iteration, models are estimated on the sample 1968Q2–1985Q2 and then used to produce forecasts for the period 1986Q2–1990Q2, etc.).

3.2.1. Direct forecasts: ARDL models

The first out-of-sample evaluation procedure is based on direct forecasts produced by bivariate ARDL models as in (1), in which the lag lengths p and q are selected recursively using the BIC. Let $\widehat{y}^h_{i,t+h|t}$ denote the direct out-of-sample forecasts of y^h_{t+h} made by the model with the ith predictor, estimated on data up to time t:

$$\widehat{y}^h_{i,t+h|t} = \widehat{\beta}^{(t)}_0 + \widehat{\beta}^{(t)}_1(L)y_t + \widehat{\beta}^{(t)}_{2,i}(L)x_{i,t} \tag{2}$$

and let $\widehat{u}_{i,t+h|t} = y^h_{t+h} - \widehat{y}^h_{i,t+h|t}$ denote the forecast error incurred by the model at time $t+h$. Each ith predictor is ranked based on the following MSFE:

$$MSFE_{i,h} = \frac{1}{T_1 - h - T_0 + 1}\sum_{t=T_0}^{T_1-h}\left(\widehat{u}_{i,t+h|t}\right)^2 \tag{3}$$

where T_0 is the end date of the shortest sample, and $T_1 - h$ is the end date of the longest sample.

Table 4 reports the MSFEs of the ARDL models relative to an AR model, used as a benchmark.[9] The log CAPR ratio is by far the best predictor over the longest horizon ($h = 20$) and the second-best predictor for both $h = 4$ and $h = 12$. The NNBLI ratio is the most effective predictor for $h = 4, 12$ and the second-best predictor at a five-year horizon. Forecast gains over the benchmark are substantial: the log CAPR ratio and the NNBLI ratio achieve MSFE values as low as 0.35 (log CAPR for $h = 20$) and 0.40 (NNBLI for $h = 12$). For $h = 4$, NNBLI has a relative MSFE of 0.55. Such results are all the more remarkable if one considers that only two variables have MSFE values lower than 0.8 for $h = 4$, and only three variables for $h = 12$ and $h = 28$. The absolute root mean square forecast error (RMSFE) also helps us appreciate the forecast performance of CAPR and NNBLI. In terms of 20-quarter cumulative GDP growth, the RMSFE of CAPR is 3.32%, corresponding to an average annual error of 0.66 percentage points of GDP growth for five years. For $h = 12$, the RMSFE of NNBLI is 2.50%, corresponding to an average annual error of 0.83%.

Other top performers include the unfilled orders for non-durable goods (*AMDMUOx*; for $h = 4, 12$), the OECD composite leading indicator (*cli*; for $h = 4, 12, 20$), the unadjusted price–rent ratio (*pr*; for $h = 4, 20$), industrial production of consumer goods (*IPCONGD*; for $h = 4, 12, 20$), the ratio of household net worth to disposable income (*NWPIx*; for $h = 4, 12$) and the inventories to sales ratio (*ISRATIOx*; for $h = 20$).

3.2.2. Iterated forecasts: VAR models

The second out-of-sample approach uses VAR models to compute multi-step-ahead iterated forecasts of GDP. The predictors are evaluated using models that include only real GDP growth and one predictor at a time. The ith VAR can be written as:

$$\widetilde{y}^{(i)}_t = a^{(i)}_0 + \sum_{j=1}^{p} B^{(i)}_j \widetilde{y}^{(i)}_{t-j} + \varepsilon^{(i)}_t \tag{4}$$

given the unbalanced nature of the dataset, the actual starting date of the sample will adjust to the availability of the time series used for estimation. In Sections 3.2.1 and 3.2.2, which consider one predictor at a time, all predictors are used regardless of the length of the respective time series. Therefore, when interpreting the results it should be kept in mind that a fraction of the predictors have fewer observations available for estimation than others. To ensure that the MSFE is computed over the same timespan for all predictors, only variables that have sufficient data to produce forecasts for 1990Q1 should be considered. However, even if the variables with an insufficient number of observations are included in the evaluation, none of them ranks among the best-performing predictors reported in Sections 3.2.1 and 3.2.2. The high-dimensional models presented in Section 3.2.3 exclude from estimation those variables (21 in total) whose time series start after 1967Q1 (as they would lead to discard observations for all other regressors), while the 14 predictors with the shortest time series (which do not have enough data to make direct forecasts for as early as 1990Q1, at least over the longest horizon) are excluded from forecast combinations.

[9] For the AR model, we consider both direct forecasts, i.e., produced by model (1) without the terms associated with x_i (and with the lag length p selected recursively by the BIC), and iterated forecasts, i.e., using specification (4), below. Since iterated forecasts are generally more accurate, we use these as our benchmark to calculate the relative MSFEs of all models. Thus, the MSFEs of ARDL (direct) forecasts and VAR (iterated) forecasts are fully comparable throughout the paper. The root mean square forecast error of the benchmark AR is 0.0174 for $h = 4$, 0.0397 for $h = 12$, and 0.0561 for $h = 20$.

Table 4
Direct forecasts by single-predictor ARDL models: Mean squared forecast errors.

	$h = 4$		$h = 12$		$h = 20$	
1	NNBTILQ027SBDIx	0.55	NNBTILQ027SBDIx	0.40	capr	0.35
2	capr	0.74	capr	0.63	NNBTILQ027SBDIx	0.53
3	AMDMUOx	0.84	cli	0.76	pr	0.56
4	pr	0.86	AMDMUOx	0.95	ISRATIOx	0.95
5	cli	0.87	NWPIx	1.01	cli	0.96
6	IPCONGD	0.88	cred_gdp	1.05	LIABPIx	1.06
7	NWPIx	0.89	UNRATESTx	1.05	UNRATESTx	1.06
8	IPNCONGD	0.89	IPCONGD	1.06	AHETPIx	1.10
9	PCECC96	0.91	FEDFUNDS	1.07	MORTGAGE30US	1.11
10	prfi_gdp	0.91	B021RE1Q156NBEA	1.07	IPCONGD	1.12

Notes: The table shows the mean squared forecast errors (MSFEs) for the h-quarter (cumulative) GDP growth rate, relative to an AR model. All models are estimated on recursive windows (shortest sample 1968Q2–1985Q1, longest sample 1968Q2–2016Q4), and MSFEs are calculated over the period 1990Q1–2017Q4. Please refer to Tables 1 and 2 for a description of the variables.

Table 5
Iterated forecasts by bivariate VAR models: Mean squared forecast errors.

	$h = 4$		$h = 12$		$h = 20$	
1	NNBTILQ027SBDIx	0.55	NNBTILQ027SBDIx	0.34	capr	0.40
2	capr	0.71	capr	0.53	NNBTILQ027SBDIx	0.48
3	pr	0.76	pr	0.69	pr	0.53
4	AMDMUOx	0.80	cli	0.71	HNOREMQ027Sx	0.80
5	TLBSNNCBBDIx	0.83	AMDMUOx	0.75	AMDMUOx	0.81
6	prfi_gdp	0.84	UNRATESTx	0.84	UNRATESTx	0.81
7	cli	0.87	USMINE	0.86	ISRATIOx	0.81
8	S&P: div. yield	0.88	IPNCONGD	0.86	B021RE1Q156NBEA	0.82
9	IPCONGD	0.88	B021RE1Q156NBEA	0.87	cli	0.83
10	cape	0.89	GS1TB3Mx	0.90	REVOLSLx	0.84

Notes: The table shows the mean squared forecast errors (MSFEs) for h-quarter-ahead GDP, relative to an AR model. All models are estimated on recursive windows (shortest sample 1968Q2–1985Q1, longest sample 1968Q2–2016Q4), and MSFEs are calculated over the period 1990Q1–2017Q4. Please refer to Tables 1 and 2 for a description of the variables.

where $\widetilde{y}_t^{(i)}$ is the vector containing the year-on-year growth rate of real GDP and the ith predictor at time t, $a_0^{(i)}$ is a 2×1 vector of constants, $B_j^{(i)}$ is a 2×2 matrix of coefficients, $\forall j = 1, \ldots, p$, and $\varepsilon_t^{(i)}$ is a 2×1 vector of error terms. The lag length p is recursively selected by the BIC for the whole system and the maximum length is again fixed at 5.

Predictors are ranked based on the performance of the VARs in terms of forecasts of the h-period-ahead GDP level:

$$MSFE_{GDP}^{(i,h)} = \frac{1}{T_1 - h - T_0 + 1} \times \sum_{t=T_0}^{T_1-h} \left(\ln(GDP_{t+h}) - \ln(\widehat{GDP}_{t+h|t}^{(i)}) \right)^2 \qquad (5)$$

where $\ln(\widehat{GDP}_{t+h|t}^{(i)})$ is the forecast of the log GDP level for period $t + h$ obtained from model i by cumulating the growth rate forecasts over time.

Table 5 reports the MSFEs of the best 10 VAR models for each horizon, relative to the benchmark AR. The top positions remain largely unchanged with respect to Table 4. Once again, the log CAPR is the best predictor for $h = 20$ and ranks second for $h = 4, 12$. The NNBLI ratio is still the most useful predictor for $h = 4, 12$.

Just as in Table 4, the forecast gains provided by CAPR and NNBLI over the benchmark AR are substantial for every h. In particular, for $h = 12$, the relative MSFE of NNBLI is 0.34, corresponding to an absolute RMSFE of 2.31% (an annual average of 0.77 percentage points of GDP growth). For $h = 20$, the log CAPR gives a relative MSFE of 0.40 and an absolute RMSFE of 3.56% (annual average: 0.71%). For $h = 4$, the relative MSFE of NNBLI is 0.55, corresponding to an absolute RMSFE of 1.3%.[10]

As for the other top predictors, private residential fixed investment (as a share of GDP) once again ranks highly over the shorter horizon of four quarters, while the OECD composite leading indicator (*cli*) and the unfilled orders for durable goods (*AMDMUOx*) rank highly over all horizons.

3.2.3. Comparison with high-dimensional forecasting models and forecast combinations

Finally, we assess whether the forecasts produced by the best single-predictor models are outperformed by those produced by methods that pool all available information. We consider three models that use all predictors at the estimation stage: (i) a large Bayesian VAR (LBVAR) using the approach by Bańbura et al. (2010); (ii) a LASSO VAR, i.e., a VAR estimated using a LASSO penalty (Hsu et al., 2008); and (iii) a factor model using principal components (a VAR model for GDP growth and a set of principal components extracted from all predictors). The three

[10] We also checked that the predictive gains provided by CAPR and NNBLI compared to the AR model were statistically significant at all horizons, using the ENC-F test by Clark and McCracken (2001), which is suitable for comparing nested models.

models reflect different approaches to the problem of high dimensionality. The LBVAR approach retains all the available predictors in the forecasting model, applying a shrinkage method that does not restrict any coefficient to be exactly zero. The LASSO VAR performs variable selection by setting a subset of coefficients exactly to zero. The principal component approach reduces the dimension of the model by summarizing the dataset of predictors into a small number of factors. We estimate the LBVAR and LASSO VAR using grids of values for the shrinkage/penalty parameters and report the best results for each forecast horizon.[11] As for the factor VAR, for each horizon we select the number of factors (between one and six) that gives the best forecasts. For all models, we produce iterated forecasts of GDP using the same recursive-window scheme as before. As in Eq. (4), we include the four-quarter growth rate of GDP in the models. To get a sense of the degree of sparsity in the LASSO VAR model, Fig. 4 depicts the regressors selected by LASSO (in the columns) for all equations of the model (in the rows).

We also consider information pooling at the forecasting stage, by calculating forecast combinations. Forecast combinations find widespread application in the forecasting literature (Chauvet and Potter 2013, Elliott and Timmermann 2016) on the grounds that individual models are likely to be misspecified and that combining forecasts from different models should increase efficiency compared to individual forecasts. Measuring the performance of combined forecasts helps us give a sense of how useful it is in practice to establish rankings of predictors: if combined forecasts turn out to outperform forecasts from every individual model, then the information contained in poorly ranking predictors should not be discarded. We combine the forecasts produced by single-predictor VAR models using two different weighting schemes. Let w_i denote the weight assigned to model i, and M the number of models to be combined. The simplest approach consists in using an equal-weighted average of the forecasts; i.e., $w_i^{equal} = 1/M$. There is ample empirical evidence that equal weighting performs well for point forecasts (e.g., Stock and Watson 2003) and often outperforms more sophisticated weighting strategies (Elliott and Timmermann 2016, Smith and Wallis 2009). The second approach is Bayesian model averaging (BMA). In particular, since the value of the BIC for model i provides an asymptotic approximation to its marginal likelihood, the BMA weights are approximated by $w_i^{bma} = \exp(-0.5BIC_i)/[\sum_{i=j}^{M} \exp(-0.5BIC_j)]$. The weights are computed recursively across estimation windows.

In addition, we report the performance of forecasts produced by a major forecasting institution, the International Monetary Fund (IMF).[12,13]

Table 6 reports the relative MSFEs of all the alternative forecasting models/methods. No model or forecast combination scheme outperforms the single-predictor models using CAPR and NNBLI at any forecast horizon. Also, unlike the small models using the best financial-cycle predictors, most models/combinations in Table 6 provide their best results relative to the benchmark AR at the shorter (four-quarter) horizon, with the exceptions of the LBVAR and the factor model.

Table 7 reports the p-values of the Diebold and Mariano (1995) (DM) test of equal forecast accuracy,[14] making comparisons on a pairwise basis between the best-performing forecasts from Tables 4–5 and the forecasts by high-dimensional models/combinations considered in this section. Under the null hypothesis, competing forecasts provide equal MSFEs, while under the alternative hypothesis, forecasts by the best predictor have lower MSFEs. At the 10% significance level, the null is generally rejected for all horizons, except for forecast combinations and the LBVAR with $h = 4$ (p-values are only slightly higher than 10% in these cases). Overall, the DM test results corroborate our findings on the predictive importance of the CAPR and NNBLI ratios. In Section 4.6, we provide results from another test of forecast accuracy, the Giacomini and Rossi (2010) test, taking into account forecast instabilities.

Fig. 3 provides a focus on the Great Recession, to assess the ability of the models presented in this section to predict it. In particular, the figure compares the pseudo-out-of-sample iterated forecasts of GDP produced by alternative models in 2007Q2, with a horizon of five

[11] The LBVAR model is estimated using a grid of possible values for the shrinkage parameters λ and τ, based on notation by Bańbura et al. (2010), where λ is an inverse measure of the tightness of the prior on the VAR coefficients, and τ determines the tightness of the additional prior on the sum of coefficients. In particular, the values considered for λ lie in the range [0.0001, 0.1], while τ can assume values of 10 or 100. We then select the values that provide the best forecast performance. The priors on model parameters are set following Alessandri and Mumtaz (2017). Forecasts are produced using the posterior means of the parameters. As for the LASSO VAR, using λ to denote the penalty parameter of the LASSO estimator, we consider $\lambda = 0.00025, 0.0005, 0.00075, 0.001, 0.00125, 0.0015$ (values outside of this range produce worse results). Again, we select the value that provides the best forecast performance. The maximum lag length is five quarters for the LBVAR and the factor model. To prevent computational problems, we consider a lag length of one for the LASSO VAR.

[12] The IMF forecasts are particularly suitable for comparisons in this context, as they cover horizons from one to five years. In the case of other major forecasters, comparisons would necessarily have limitations in terms of horizons. For instance, the Fed's Greenbook forecasts cover a maximum horizon of two years. The OECD publishes annual forecasts for the following year and long-term projections in 10-year steps. The Survey of Professional Forecasters (SPF) includes quarterly forecasts up to four quarters ahead, annual forecasts for the next three years (albeit only starting from the 2009Q2 survey), and 10-year annual average forecasts. However, as Pain et al. (2014) point out, "the profile and magnitude of the errors in the GDP growth projections [over 2007–2012] of other international organizations and consensus forecasts are strikingly similar".

[13] The IMF's *World Economic Outlook* provides forecasts in the form of annual growth rates up to five years ahead. We convert them into 4-quarter-ahead, 12-quarter-ahead, and 20-quarter-ahead forecasts of the GDP level using the following approach: (i) we take the spring issues of the *World Economic Outlook*, (ii) we consider the last quarter of the year prior to each issue as the starting value for forecasting, (iii) we apply the annual forecast growth rates to the starting value to compute forecasts of the GDP level, and (iv) we assign each resulting value to the last quarter of the relevant forecast year. For example, the annual forecast growth rate for 1990 published in the Spring 1990 issue is used to compute the 4-quarter-ahead forecast for 1990Q4 based on historical data up to 1989Q4, the annual rates up to 1992 are used to compute the 12-quarter-ahead forecast for 1992Q4 based on data up to 1989Q4, and so on.

[14] We use the test correction proposed by Harvey et al. (1997).

Table 6
High-dimensional models and forecast combinations: Mean squared forecast errors.

Forecasting model/method	$h = 4$	$h = 12$	$h = 20$
LBVAR	0.82	0.90	0.75
LASSO VAR	0.87	0.90	0.90
Factor model	0.91	0.96	0.79
Forecast combination (equal weights)	0.93	0.99	1.00
Forecast combination (BMA weights)	0.96	0.99	1.00
IMF	0.76	1.00	1.33

Notes: The table shows the MSFEs for the h-quarter-ahead cumulative GDP growth rate, relative to the benchmark AR model, over the period 1990Q1–2017Q4. The forecasting models/methods considered are a large Bayesian VAR (LBVAR); a VAR estimated using a LASSO penalty (LASSO VAR); a factor model; combinations of forecasts from one-predictor VAR models, using equal weights and Bayesian model averaging (BMA) weights; and forecasts by the International Monetary Fund (IMF).

Table 7
Diebold–Mariano test: Best predictors vs. high-dimensional models.

Forecasting model/method	$h = 4$	$h = 12$	$h = 20$
LBVAR	0.12	0.07	0.04
LASSO VAR	0.07	0.07	0.09
Factor model	0.04	0.01	0.09
Forecast combination (equal weights)	0.12	0.10	0.11
Forecast combination (BMA weights)	0.10	0.10	0.11
IMF	0.01	0.09	0.00

Notes: For each horizon h, the table reports the p-values of the Diebold–Mariano test of equal MSFEs comparing the GDP forecasts using the best predictor from Tables 4–5 and those by high-dimensional models/methods from Table 6. Under the null hypothesis, competing forecasts have equal MSFEs. Under the alternative hypothesis, forecasts by the best predictor have lower MSFEs. All MSFEs are computed over the period 1990Q1–2017Q4.

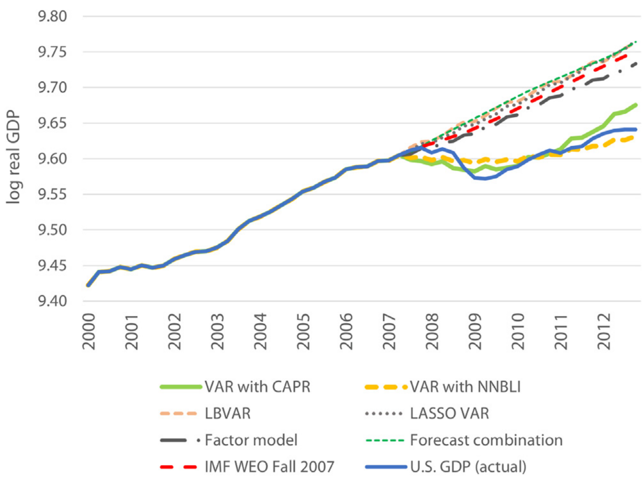

Fig. 3. Forecasting the great recession: A comparison of models.
Notes: The figure shows the pseudo-out-of-sample forecasts of U.S. GDP over a five-year horizon produced in 2007Q2 by alternative models: bivariate VAR with CAPR or NNBLI, a large Bayesian VAR (LBVAR), LASSO VAR, and a factor model. The figure also reports forecast combinations using equal weights and the forecasts contained in the Fall 2007 *World Economic Outlook* by the IMF. All models are estimated over the sample 1967Q1–2007Q2.

years. All models are estimated over the sample 1968Q2–2007Q2, i.e., using only observations prior to the beginning of the Global Financial Crisis. The figure also plots forecast combinations using equal weights and the forecasts published by the IMF in its Fall 2007 *World Economic Outlook*. As shown in the figure, small models based on CAPR and NNBLI are much more effective at forecasting the Great Recession and the subsequent slow recovery than larger models.

4. Extensions and robustness checks

In this section, we present a number of extensions to our main results as well as many robustness checks.

First, we investigate whether the predictive power of CAPR and NNBLI is limited to recession periods, perhaps characterized by financial tightening. The recent literature on growth at risk (Adrian et al. 2019, 2022) has found that the NFCI, capturing general financial conditions, is

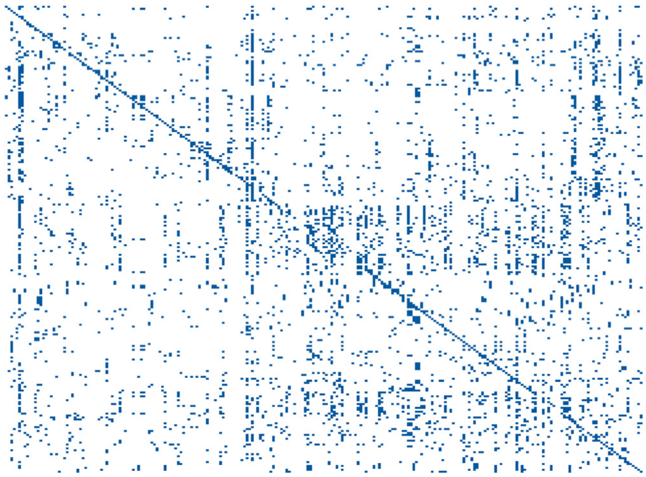

Fig. 4. High-dimensional models: Sparsity in the LASSO VAR.
Notes: The figure provides a visual summary of variable selection in the LASSO VAR model, estimated over the sample 1968Q2–2016Q4. The figure depicts a table in which each row corresponds to an equation in the large VAR and each column identifies a right-hand-side variable (predictor). Colored cells identify predictors with non-zero coefficients.

useful to forecast tail risks to GDP, but is a weak predictor of the rest of the conditional GDP distribution. Following this literature, in Section 4.1 we estimate quantile regressions for GDP growth and show that the strong predictive relationship of CAPR and NNBLI with economic activity is stable across lower and upper percentiles of the GDP distribution (especially for CAPR). That is, it appears to be present during both expansion and recession periods, unlike for the NFCI.

Second, we assess the ability of variables to specifically forecast the year-on-year growth rate at quarter $t+h$, instead of the growth rate over h quarters (i.e., cumulative growth, and hence the h-quarter-ahead GDP level) considered so far, and show that the best forecasts are still provided by VAR models using CAPR and NNBLI.

Next, we address the critical issue of forecast instabilities. A large body of literature has shown that the forecasting power of variables is generally unstable over time, i.e., that some predictors provide good forecast performance but only in specific periods (Clements and Hendry 2006, Rossi 2021, Stock and Watson 2003, among others). More specifically, we address the questions: To what extent are the main results determined by the Global Financial Crisis and Great Recession period? Would it have been possible to identify CAPR and NNBLI as top predictors in previous periods? Also, some literature on parameter instabilities finds that time-varying-parameter (TVP) models, in particular TVP-VARs, tend to produce better forecasts than models with time-invariant parameters (e.g., D'Agostino et al. 2013, Koop and Korobilis 2013). We assess the robustness of our main results to forecast and parameter instabilities in several ways. In Section 4.3, we repeat the out-of-sample evaluation of Section 3.2, estimating the models on rolling windows, which accommodate structural breaks in parameters. In Section 4.4, we consider forecasts produced by TVP-VARs. In Section 4.5, we conduct in-sample and out-of-sample evaluations over sub-periods. In Section 4.6, we use the Giacomini and Rossi (2010) test to account for instabilities when comparing forecast performance.

Finally, in Section 4.7, we compare different variable-selection methods, to check whether they select the same predictors, in particular CAPR and NNBLI.

To facilitate reading, in all the tables reported in the following sections, we highlight CAPR and NNBLI using bold text. Also, in the interest of space, we place several tables and figures in the Online Appendix.

4.1. Quantile regressions

First, we explore the predictive ability of CAPR and NNBLI at different quantiles of the conditional distribution of GDP growth. To this aim, we first estimate quantile regressions and evaluate predictors using the local goodness-of-fit measure at the specific quantiles introduced by Koenker and Machado (1999). The estimated models are the same ARDL models as in Eq. (1). Table 8 reports results for the 10th and 90th percentiles of the h-period GDP growth (very similar results are obtained for the 5th and 95th percentiles), listing only the top five predictors in each case to save space. CAPR is among the most powerful predictors for both quantiles, while NNBLI is a top predictor for the upper quantile but not for the lower (although its results are not very far from the best ones for $h=20$ and $h=12$).

Interestingly, for the lower quantile and for $h=4$, powerful predictors include a term spread ($T5YFFM$), which has long been used as a leading indicator of recessions (Chauvet and Potter 2013, Stock and Watson 2003), and the composite National Financial Conditions Index (NFCI) (sixth in the ranking, with a pseudo-R^2 of 0.37, and hence not reported in Table 8), which are both weaker predictors of the upper quantile. This is consistent with the growth-at-risk literature (Adrian et al., 2019), which has previously shown the NFCI to provide valuable predictive content only for the left tail of the conditional GDP distribution. Reichlin et al. (2020) further investigate this

Table 8
Goodness of fit of quantile regressions for GDP growth.

	$h = 4$		$h = 12$		$h = 20$	
			Quantile: 10%			
1	prfi_gdp	0.43	AMDMUOx	0.51	pr	0.53
2	PRFIx	0.42	**capr**	0.45	ISRATIOx	0.53
3	T5YFFM	0.39	pr	0.43	**capr**	0.51
4	cli	0.38	prfi_gdp	0.43	mortg_inc	0.48
5	PERMITMW	0.37	ISRATIOx	0.42	LIABPIx	0.48
			Quantile: 90%			
1	cred_gd	0.38	**capr**	0.50	**capr**	0.51
2	**capr**	0.35	**NNBTILQ027SBDIx**	0.43	**NNBTILQ027SBDIx**	0.48
3	**NNBTILQ027SBDIx**	0.35	pr	0.41	pr	0.44
4	LIABPIx	0.35	cli	0.40	cred_gd	0.40
5	mortg_inc	0.35	cred_gd	0.40	LIABPIx	0.34

Notes: The table reports the local goodness-of-fit measure (pseudo-R^2) by Koenker and Machado (1999) for quantile regressions. For each h, the dependent variable is GDP growth over h quarters. All models are estimated using data from 1974Q1 to 2017Q4. Please refer to Table 1 and McCracken and Ng (2021) for a description of the variables.

result by distinguishing between different components of the NFCI. They find that price variables, such as credit spreads, actually provide limited information on growth vulnerability, whereas non-financial leverage (which is related to the NNBLI ratio) has provided useful early warnings for GDP in the Global Financial Crisis.

To better characterize our findings and to further connect to the literature on growth vulnerabilities, we next consider the coefficients of the quantile regressions of GDP using (alternatively) the NFCI, CAPR, and NNBLI as predictors, and we check if and how these coefficients vary across quantiles of GDP. Fig. 5 displays the coefficients for different quantiles of GDP from 5% to 95%. We report results for the simplest ARDL regressions with only an h-quarter lagged term for each predictor, i.e., Eq. (1) with $p = q = 1$. For each regression, the figure reports the point estimate of the coefficient (blue line) and the 95% confidence intervals (red lines). A dotted horizontal line indicates the value of zero. The results are quite revealing. The NFCI has negative and significant coefficients only for lower percentiles of GDP over horizons of one to three years (meaning that higher levels of financial stress predict lower GDP growth in the following periods), while coefficients are non-significant and positive in point estimates for higher percentiles and for all percentiles at the longest horizon $h = 20$. These results are broadly consistent with those by Adrian et al. (2019) and Adrian et al. (2022). Conversely, the coefficients on CAPR and NNBLI are invariably negative across all quantiles and horizons, and in general strongly significant (with a few exceptions at low quantiles), their absolute magnitude increasing with h. Thus, the relationship of CAPR and NNBLI with GDP appears stable across different parts of the GDP distribution. Overall, these results indicate that the predictive power of the two ratios is not simply related to recessions induced by financial distress, unlike for measures of general financial conditions. The results presented in the remainder of the paper confirm that CAPR and NNBLI achieve good predictive performance during both recession and expansion periods.

4.2. Forecasting year-on-year growth rate at horizon *h*

Following the Stock and Watson (2003) approach, this paper focuses on predicting GDP growth over h periods (i.e., cumulative growth) and therefore the h-period-ahead GDP level. Multi-year growth also appears of special interest from more recent and related papers dealing with forecasts of GDP over medium-term horizons (e.g., Mian et al. 2017 on household debt and GDP, or Adrian et al. 2022 on the term structure of growth at risk). As an extension to this main approach, we now evaluate forecasts of the specific growth rate of GDP at a given future quarter $t+h$, irrespective of GDP movements in previous periods. We report results on year-on-year growth rates, i.e., relative the same quarter of the previous year, but similar qualitative results were obtained for quarter-on-quarter growth rates.

Table 9 reports the best ARDL and VAR forecasts for this target variable. Of course, the results for $h = 4$ coincide with those already presented in Tables 4–5. As Table 9 shows, the best forecasts are still provided by VAR models using CAPR and NNBLI. In particular, CAPR and the unadjusted price–rent ratio produce the best forecasts for both $h = 12$ and $h = 20$, while NNBLI is the best predictor for $h = 4$ and the third-best predictor for $h = 12$. ARDL forecasts do not achieve the same level of accuracy for this target variable.

In the Online Appendix, we also report (Table A1) the relative MSFEs of the high-dimensional models/methods considered in Section 3.2.3. All models/methods perform better than the benchmark AR at short horizons, and LBVAR and LASSO VAR produce good forecasts at all horizons. Still, all models are outperformed by the best one-predictor models at all horizons.

4.3. Forecasts using rolling-window estimates

We now go back to our main goal of forecasting GDP growth over h periods, and assess the extent to which the out-of-sample evaluation of Section 3.2 is robust to estimation on different samples.

In particular, we calculate the relative MSFEs of forecasts produced by ARDL and VAR models using rolling

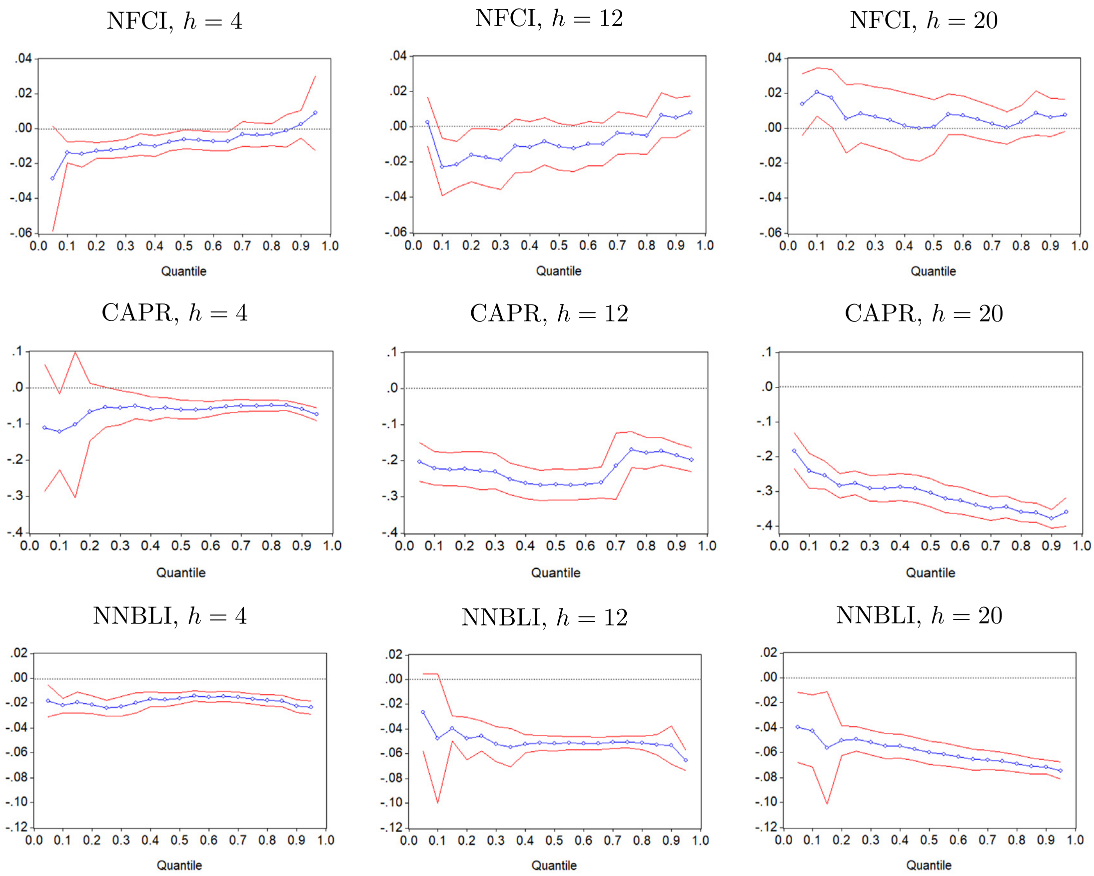

Fig. 5. Coefficients of quantile regressions of GDP.
Notes: The figure shows the estimated coefficients associated with different predictors in quantile regressions for *h*-quarter cumulative GDP growth (with 95% confidence intervals). The variable NNBLI is divided by 100,000 to make the results easier to read. Dotted horizontal lines indicate the value of zero.

Table 9
Forecasts of *h*-period-ahead year-on-year GDP growth rate: MSFEs.

	$h = 4$		$h = 12$		$h = 20$	
			ARDL			
1	**NNBTILQ027SBDIx**	0.55	AAA	0.93	AAA	0.90
2	**capr**	0.74	cli	0.93	BAA	0.92
3	AMDMUOx	0.84	gs10	0.93	TCU	0.92
4	pr	0.86	GS5	0.94	MZMREALx	0.93
5	cli	0.87	BAA	0.95	CUMFNS	0.94
			VAR			
1	**NNBTILQ027SBDIx**	0.55	**capr**	0.68	**capr**	0.71
2	**capr**	0.71	pr	0.74	pr	0.72
3	pr	0.76	**NNBTILQ027SBDIx**	0.88	ISRATIOx	0.79
4	AMDMUOx	0.80	ISRATIOx	0.89	B021RE1Q156NBEA	0.89
5	TLBSNNCBBDIx	0.83	UNRATESTx	0.90	REVOLSLx	0.93

Notes: Out-of-sample MSFEs for the *h*-quarter-ahead year-on-year GDP growth rate, relative to the benchmark AR. All models are estimated on recursive windows (shortest sample 1968Q2–1985Q1, longest sample 1968Q2–2016Q4), and MSFEs are computed over the period 1990Q1–2017Q4. Please refer to Table 1 and McCracken and Ng (2021) for a description of the variables.

windows with a fixed length, instead of the recursive windows used for the baseline results of Section 3.2. We considered rolling windows of 40, 60, and 80 quarters, and report results for the mid-size, i.e., 60 quarters (Table A2 in the Online Appendix shows the top 10 predictors and the associated MSFEs for each horizon). Thus, for instance, forecasts generated at time 2007Q2 are based on estimates obtained over the sample 1992Q3–2007Q2 instead of 1968Q2–2007Q2. The MSFE is calculated relative to the benchmark (recursive-scheme) MSFE of the

AR model from Section 3.2. As mentioned above, rolling windows generally provide less accurate top forecasts compared to recursive windows.[15]

The results are in line with those of Section 3.2, especially for NNBLI. Considering the ARDL and VAR forecasts jointly, NNBLI is the best predictor for $h = 4$ and $h = 12$ and the second-best for $h = 20$. The unadjusted price–rent ratio is the best predictor for $h = 20$, and CAPR is among the top predictors at all horizons. In the case of VAR forecasts, NNBLI is the top performer at all horizons, while CAPR is second for $h = 4$, third for $h = 20$, and fifth for $h = 12$. Other top predictors include the unfilled orders for durable goods (*AMDMUOx*), non-financial corporate business sector net worth to disposable business income (*TNWMVBSNNCBBDIx*), the OECD composite leading indicator (*cli*), and residential fixed investment (*PRIx*).

4.4. Forecasting with time-varying-parameter VARs

The results obtained using rolling-window estimates suggest that our main findings are quite robust to parameter instabilities. In this section, we go a step further in this direction. So far, we have considered models whose parameters do not vary within the estimation sample. As mentioned above, previous literature has found that time-varying-parameter VARs (TVP-VARs) produce better forecasts than models with time-invariant parameters (e.g., D'Agostino et al. 2013, Koop and Korobilis 2013). We now consider forecasts from TVP-VARs and check whether our main results are confirmed. We evaluate the TVP-VARs using the baseline recursive-window scheme.

We first consider bivariate TVP-VARs with GDP growth and one predictor at a time.[16] The upper part (Panel A) of Table 10 reports the results for the top five predictors. The MSFEs are expressed relative to the benchmark time-invariant-parameter AR model. The table broadly confirms our main findings: CAPR and NNBLI are the best two predictors at all horizons. Using TVP-VARs further improves the forecast performance of NNBLI for $h = 4$ and $h = 12$, compared to Table 5, while forecasts of both CAPR and NNBLI for $h = 20$ are worse than their time-invariant-parameter counterparts.

Next, we extend our set of high-dimensional models by considering the large TVP-VAR approach by Koop and Korobilis (2013). In particular, these authors propose dynamic model selection (DMS) and dynamic model averaging (DMA) methodologies to mix predictions from TVP models of different sizes. We implement the TVP-VAR-DMS and TVP-VAR-DMA approaches using two TVP-VARs.[17] The smaller TVP-VAR is the three-variable model used by Koop and Korobilis (2013), which includes GDP, inflation, and an interest rate (the Fed funds rate). The larger TVP-VAR includes the set of 18 predictors selected by the LASSO estimator (over the full sample) in the GDP equation of the LASSO VAR of Section 3.2.3 (as explained in Section 4.7, the list of predictors is provided in the Online Appendix). We consider lag orders from one to five and only report the best results for each forecast horizon. The lower part (Panel B) of Table 10 displays the MSFEs of TVP-VAR-DMA, which provides better forecasts than TVP-VAR-DMS in this context. The approach achieves good performance for $h = 4$, with a relative MSFE of around 0.9, close to the values obtained by high-dimensional models from Table 6 (in particular, the factor model and the LASSO VAR), whereas its forecast accuracy deteriorates at longer horizons. More importantly, it is outperformed by the best one-predictor VARs (both with time-invariant and time-varying parameters) at all horizons. This is in part consistent with Koop and Korobilis (2013), who find that small TVP-VARs tend to be preferred over larger models when forecasting GDP (while larger TVP-VARs and TVP-VAR-DMS/DMA are more useful for inflation and interest rates).

4.5. Evaluation over sub-samples

As a third check on the robustness of results to instabilities in predictive power, we evaluate predictions over sub-periods. In particular, we further address the questions: Do the main results from Sections 3.1 and 3.2 simply depend on the inclusion of the Global Financial Crisis (GFC) and Great Recession in our sample? Would it have been possible to identify CAPR and NNBLI as top predictors before the GFC?

To begin with, we calculate the R^2 of ARDL models from Section 3.1, excluding the period 2007Q3–2009Q2, i.e., the GFC and Great Recession period. We take the third quarter of 2007 as the beginning of the GFC (in particular, the month of August, when BNP Paribas stopped withdrawals from three of its hedge funds and major indicators of financial stress rose, e.g., the NFCI). The second quarter of 2009 is the end of the Great Recession, according to the NBER recession dates. The results are reported in the upper part of Table A3 in the Online Appendix. CAPR, the unadjusted price–rent (PR) ratio, and NBBLI are the top three predictors for $h = 12$ and $h = 20$, and CAPR is still among the top predictors for $h = 4$. Next, we calculate the R^2 on periods before the GFC. Part (b) of Table A3 in the Online Appendix shows the results obtained using data from 1974Q1 to 2007Q2 as in

15 The accuracy of top forecasts increases with the size of the rolling windows. However, the results for 80 quarters are similar to those for 60.

16 We estimate TVP-VARs à la Primiceri (2005), which also allow for stochastic volatility of errors. The VAR coefficients are assumed to follow a random walk process, so their best estimate for out-of-sample periods corresponds to the value estimated for the last quarter of the sample window. We estimate models using code by Gary Koop, available at https://sites.google.com/site/garykoop/, and report results based on 1000 Markov chain Monte Carlo (MCMC) replications and 1000 burn-in replications. To reduce the computational burden, we do not implement automatic lag selection as in Section 3.2, but assume a fixed lag order of two in all models. In each sample window, the prior for the TVP-VAR is given by the OLS estimates of time-invariant parameters obtained using data within that window. We also consider uninformative priors, which, however, produce less accurate top forecasts.

17 We use software code provided by Koop and Korobilis (2013). Based on their results, we set the "forgetting factors" to the value of 0.99 and the "decay factor" to 0.96 (see Koop and Korobilis 2013 for details).

Table 10
TVP-VAR forecasts: MSFEs.

	$h = 4$		$h = 12$		$h = 20$	
			Panel A: **Bivariate TVP-VAR**			
1	**NNBTILQ027SBDIx**	0.53	**NNBTILQ027SBDIx**	0.28	**capr**	0.46
2	**capr**	0.71	**capr**	0.57	NNBTILQ027SBDIx	0.50
3	pr	0.79	pr	0.73	pr	0.57
4	prfi_gdp	0.80	AMDMUOx	0.73	NWPIx	0.78
5	hpi	0.81	cli	0.79	AMDMUOx	0.80
			Panel B: **Large TVP-VAR**			
	0.91		0.98		1.14	

Notes: The table reports the MSFEs for the *h*-quarter (cumulative) GDP growth rate, relative to the benchmark AR, computed over the period 1990Q1–2017Q4, using bivariate TVP-VAR models (Panel A) and the TVP-VAR-DMA approach by Koop and Korobilis (2013) (Panel B). See Table 1 and McCracken and Ng (2021) for a description of the predictors listed in Panel A. See Section 4.4 for details on the TVP-VAR-DMA approach.

Section 3.1, while part (c) shows the results over a sample corresponding (approximately) to the Great Moderation period, 1983Q1–2007Q2. CAPR is invariably among the top predictors for $h = 12$ and $h = 20$, while NNBLI provides the best fit for $h = 12$ and the second-best fit for $h = 4$ during the Great Moderation period.

We then consider pseudo-out-of-sample forecasts by ARDL and VAR models, and split the forecast evaluation period in a pre-Crisis period (1990Q1–2007Q2), a Crisis period (2007Q3-2009Q2), and a post-Crisis period (2009Q3–2017Q4). Tables A4–A5 in the Online Appendix report the MSFEs over these different periods. Overall, the forecasting power of CAPR and NNBLI appears stable compared to the other predictors. In the pre-Crisis period, the best forecasts are provided by NNBLI at all horizons (in particular, VAR forecasts for $h = 4$ and ARDL forecasts for $h = 12$ and $h = 20$), while CAPR appears especially useful for long horizons: it is the second-best predictor for $h = 20$ (using ARDL forecasts), and among the best predictors for $h = 12$. During the Crisis period, the best forecasts for $h = 4$ are provided by CAPR and NNBLI, along with private residential fixed investment (*PRFIx*). For $h = 12$, the best forecasts are again provided by CAPR and NNBLI, along with house price growth and the ratio of private residential fixed investment to GDP. CAPR also provides the best forecasts for $h = 20$ (ARDL). In the post-Crisis period, the two ratios are again the best predictors (both ARDL and VAR) for $h = 12$ and $h = 20$, together with the unadjusted price–rent ratio, while they are relatively less powerful at the shorter horizon $h = 4$.[18]

Overall, NNBLI-based forecasts appear better for horizons of one to three years, and CAPR-based ones for three to five years. NNBLI would be unambiguously selected before the GFC for all horizons, while CAPR only in the case of the longest horizon.

Finally, we focus on the ability of the two ratios to forecast recessions. Above, Fig. 3 showed that both CAPR and NNBLI were remarkably effective in forecasting the Great Recession. In the Online Appendix (Figure A1), we assess their ability to forecast other recessions, namely the 1990–1991 recession and the 2001 recession. In particular, we consider forecasts produced by bivariate VAR models using either CAPR or NNBLI, estimated on data up to 1990Q2 or up to 2000Q4. In both cases, the two variables predict slowdowns in economic activity. NNBLI is more effective than CAPR at forecasting the 2001 recession.

4.6. Testing forecast accuracy in the presence of instabilities: The Giacomini–Rossi (2010) test

Giacomini and Rossi (2010) proposed a methodology to test the forecasting performance of competing models in a way that is robust to the presence of instabilities. As a further check, for each forecast horizon *h*, we take the best forecasts from Tables 4–5 (i.e., NNBLI-based VAR forecasts for $h = 4$ and $h = 12$ and CAPR-based ARDL forecasts for $h = 20$) and perform the Giacomini and Rossi (2010) test comparing these one-predictor forecasts with the forecasts produced by the high-dimensional approaches considered in Section 3.2.3.

Table 11 reports the values of the test statistics, along with 5% and 10% critical values.[19] The Giacomini–Rossi test results are broadly in line with the Diebold–Mariano test results from Table 7. In particular, one-predictor model forecasts for $h = 12$ and $h = 20$ significantly outperform all forecasts from high-dimensional models at the 5% level of significance and the IMF forecasts at the 10% level, while forecast improvements are typically not statistically significant at the shorter horizon $h = 4$, except relative to the IMF forecasts at the 10% level.

In the Online Appendix (Figure A2), we provide a plot of the entire rolling sequence of out-of-sample loss differences between competing forecasts, along with the 5% critical values of the test, from 1999Q4 to 2017Q4. NNBLI-based forecasts for $h = 12$ and CAPR-based forecasts for

[18] When the MSFEs were calculated over a sample that only excluded the Crisis period (not reported in the interest of space), CAPR was still the best predictor in the case of both ARDL and VAR forecasts for $h = 20$, NNBLI was the best predictor and CAPR the second-best for $h = 12$, and NNBLI was the best predictor (using both ARDL and VAR) for $h = 4$.

[19] The Giacomini–Rossi test is based on forecast loss differences between competing forecasts, calculated over rolling windows of out-of-sample observations. In our case, the time series of forecast errors run from 1990Q1 to 2017Q4 (112 quarters). We calculate loss (squared-error) differences on rolling widows of 40 quarters (so the first window is 1990Q1–1999Q4, the second is 1990Q2–2000Q1, and so on), but similar results were obtained using windows of 60 quarters. We calculate the test statistics using Newey–West heteroskedasticity and autocorrelation consistent (HAC) estimators of the long-run variances of the loss differences, with a bandwidth of two lags, based on the results from AR regressions, indicating that two lags generally capture all the significant autocorrelation in loss differences.

Table 11
Comparing forecast accuracy in the presence of instabilities: The Giacomini–Rossi (2010) test.

	LBVAR	LASSO VAR	Factor	Combin.	IMF	10% c.v.	5% c.v.
NNBLI, $h = 4$	2.424	2.126	2.364	2.058	2.630	2.626	2.890
NNBLI, $h = 12$	3.185	5.518	6.639	4.138	2.830	2.626	2.890
CAPR, $h = 20$	6.404	4.346	3.511	4.661	2.843	2.626	2.890

Notes: The table reports the Giacomini–Rossi (2010) test statistics, along with 5% and 10% critical values (in the last two columns). For each horizon h, the test compares the best forecasts from one-predictor models with the forecasts produced by the high-dimensional approaches considered in Section 3.2.3. The test is calculated using forecast errors in the period 1990Q1–2017Q4. Loss differences between competing models are calculated on rolling windows of 40 quarters. Under the null hypothesis, competing models have equal predictive ability. Under the alternative hypothesis, one-predictor models outperform high-dimensional approaches.

$h = 20$ tend to significantly outperform forecasts by high-dimensional models/methods both before and after the Global Financial Crisis and Great Recession period.

4.7. Comparing variable-selection methods

In Section 3.1, predictors were considered one at a time, and CAPR and NNBLI stood out in the evaluation based on R^2. As a further robustness check, we now consider widely used variable-selection methods, i.e., approaches for selecting subsets of predictors after pooling all the available data, to check whether they also select the same two predictors.[20] We consider three different approaches: LASSO-based variable selection (Tibshirani, 1996); Bayesian variable selection using a popular shrinkage prior, namely the horseshoe prior by Carvalho et al. (2010); and variable selection based on the random forest methodology (Genuer et al., 2010), which is popular in machine learning. We apply these methods to ARDL models of GDP growth, considering all predictors at the same time. In the case of the LASSO, we also report the list of variables selected for the GDP equation of the LASSO VAR introduced in Section 3.2, estimated on the largest sample.[21]

LASSO applied to ARDL selects NNBLI for all values of h and CAPR for $h = 12$ and $h = 20$. The horseshoe prior selects NNBLI for all horizons, but not CAPR. In the case of random forests, predictors are actually ranked (not just selected) on the basis of a measure of variable importance, capturing their contribution to the goodness of fit.[22] This approach selects CAPR for all horizons, indicating it as the best predictor for both $h = 12$ and $h = 20$, while the unadjusted price–rent ratio is the best predictor for $h = 4$. NNBLI is selected for $h = 4$ and $h = 12$. Finally, the LASSO VAR selects NNBLI but not CAPR.[23] Table A6 in the Online Appendix reports the complete list of variables selected by the different methods.

Overall, these results confirm the importance of NNBLI and CAPR. In particular, NNBLI is the only variable that is selected by all methods considered here, and for all horizons (except $h = 20$ in the random forest). As above, CAPR appears to be especially useful for horizons of three to five years.

4.8. Forecast evaluation using real-time data

Lastly, we evaluate forecasts using historical data vintages, to check that our main results are not simply determined by the use of revised data. A fully real-time analysis is not feasible, due to limited availability of data vintages. In particular, the only vintage available for the FRED-QD dataset before 2018 is the beta version of November 2015. However, a large subset of variables is also included in the smaller FRED-MD dataset (McCracken & Ng, 2016), for which a complete series of vintages is provided starting from September 1999. Therefore, we produce and evaluate real-time forecasts based on the complete set of vintages of FRED-MD (128 variables), using the 1999Q4 vintage for earlier periods.[24] We also check that our main results are confirmed when using the 2015-11 vintage of the complete FRED-QD dataset.

For GDP, which is used both as the target variable and as a regressor in the ARDL and VAR models, we retrieve the complete set of real-time vintages starting from December 1991 from the St. Louis Fed's Archival FRED (ALFRED) database. We use real-time GDP data to produce forecasts, and final (revised) data to evaluate them.

For CAPR, data from Davis et al. (2008) on prices and rents for the aggregate stock of owner-occupied housing

20 The pool of variables includes all FRED-QD variables, transformed as indicated by McCracken and Ng (2021), and the additional variables from Table 1.

21 For LASSO ARDL models, we first considered tuning the penalty parameter through a 10-fold cross-validation procedure, i.e., finding the value that minimizes the cross-validated MSE. However, this resulted in a large set of selected variables for all h, so for the ease of exposition – and to more clearly highlight the importance of CAPR and NNBLI – we report results using a stricter penalty ($\lambda = 0.005$), i.e., further shrinking the list of selected predictors. For LASSO VAR, we use the same value of the penalty parameter as in Section 3.2, i.e., the one that provides the best forecasts of GDP. In the case of the horseshoe prior, variables are selected when the 90% credible interval for their coefficient does not include zero. In more detail, the implemented horseshoe prior is based on a Cauchy prior truncated to $[1/k, 1]$, where k denotes the total number of predictors, and a Jeffreys prior for the error variance. The results are obtained using 5000 Markov chain Monte Carlo (MCMC) samples and 2000 burn-in samples.

22 Specifically, the importance of each variable is measured as the increase in the out-of-bag MSE resulting from a random permutation of that variable, averaged over all trees in the random forest. Only the best variables are kept, based on the thresholding strategy proposed by Genuer et al. (2010). The results are based on 5000 trees.

23 It should be noted, however, that since VAR estimation by construction minimizes the one-step-ahead mean squared error, the LASSO VAR approach may simply fail to recognize CAPR as a top predictor because this variable exhibits strong predictive power at long horizons. Moreover, each of the variables in the GDP equation has in turn its own equation in the LASSO VAR, so many more variables are indirectly involved in GDP forecasts in this case, as shown by Fig. 4, introduced above.

24 In each quarter, we use the vintage released in the last month of the quarter. We use 1999Q4 as the first vintage, instead of 1999Q3, because many variables are not available in the 1999Q3 vintage.

are not available in real time. However, similar values of the ratio are obtained using the Case–Shiller house price index and the Owners' Equivalent Rent of Residences in U.S. City Average by the U.S. Bureau of Labor Statistics, for which historical vintages are available in ALFRED, starting from April 2011 for rents and from November 2014 for house prices. Moreover, for these series, data revisions are generally negligible, so that CAPR data can be considered approximately real-time even before 2011 (to get real values, we divide by the non-seasonally-adjusted consumer price index, which is not revised over time).[25] For NNBLI, we can only construct real-time vintages from 2010Q1, using raw data on non-corporate liabilities and business disposable income available in ALFRED (see McCracken and Ng 2021 for details on data elaboration).

For the additional variables in Table 1, we collect real-time data from ALFRED when available. In general, the earliest vintages in ALFRED are only available after 2010. For the composite leading indicator, real-time vintages are provided by the OECD, starting in May 2001.

Table A7 in the Online Appendix reports the results for ARDL and VAR forecasts using the available data vintages. CAPR and NNBLI are confirmed as the top predictors at all horizons, along with the unadjusted price–rent ratio. Although these results must be taken with caution, especially for NNBLI, they still suggest that the predictive power of the two ratios is quite robust to the use of alternative historical vintages.

5. Insights from macro-finance theory

The results in the previous sections reveal that the CAPR and NNBLI ratios have a robust negative relationship with economic growth over the subsequent years. But theoretically, what justifies this predictive relationship? Intuitively, since the two ratios combine information on financial conditions and economic fundamentals, they appear useful for signaling financial vulnerabilities. However, a discussion of the macro-finance theory provides deeper insights on the mechanisms linking these ratios to aggregate activity.

The role of firms' debt as a key driver of business-cycle fluctuations has long been recognized by macroeconomic theory (e.g., Bernanke et al. 1999, Kiyotaki and Moore 1997). In particular, the financial accelerator model by Bernanke et al. (1999), arguably the most influential framework for studying the macroeconomic effects of business debt (Brunnermeier & Krishnamurthy, 2020), hinges on the relationship between debt and firms' net worth, which is determined by business income and asset values. In the presence of credit market frictions, such as agency costs, lower (higher) net worth relative to debt leads to higher (lower) costs of financing, thereby decreasing (increasing) borrowing, investment, and production. As highlighted by Brunnermeier and Krishnamurthy (2020), the financial accelerator framework is built on a corporate finance model (namely, an entrepreneur-manager firm model) that is most suitable for small firms. Since the standard organizational models of small firms are non-corporate ones, such as sole proprietorships and partnerships, this theoretical framework provides a strong rationale for the informative role of the debt-to-income ratio of non-corporate businesses, i.e., NNBLI. Moreover, in models like Bernanke et al. (1999) and Kiyotaki and Moore (1997), durable goods such as houses are used as collateral for lending, so fluctuations in their prices represent an important amplification mechanism, affecting the availability of credit and production. Since collateralized borrowing is especially important for small businesses (e.g., Banerjee and Blickle 2021), this framework also provides a straightforward theoretical link between the predictive power of NNBLI and that of housing market valuation metrics.[26]

In recent years, the macro-finance literature has further expanded on the macroeconomic role of housing, as documented in the surveys by Duca et al. (2021), Piazzesi and Schneider (2016), and Davis and Van Nieuwerburgh (2015), highlighting the mutual interactions between house prices and debt. Given the key role of housing in collateralized lending, changes in house prices affect economic activity through their impact on credit to households and firms, the banking sector, and the broader financial system. They also affect output through direct wealth effects. These effects are typically strong compared to those generated by other types of assets (e.g., stocks), due to the high marginal propensity to consume out of housing wealth, which is held to a large extent by indebted, low-net-worth households (e.g., Mian and Sufi 2015). In good times, increases in house prices lead to relaxed credit constraints, more investment and consumption, more production, and thus further upward pressures on prices. Conversely, house price declines lead to tightened credit conditions, deleveraging, and downturns. Against this backdrop, a series of recent macro models with a housing sector (e.g., Favilukis et al. 2017, Justiniano et al. 2019, Sommer et al. 2013) explicitly investigate the relationship between aggregate economic activity and the price–rent ratio, which is indicated by the asset pricing theory as a key valuation metric for housing, capturing the relationship between market values and fundamentals (e.g., Campbell et al. 2009, Kishor and Morley 2015).[27] Overall, the literature suggests that, because of the role of housing in collateralized credit and the

[25] Since the FRED rent time series starts in 1983, to calculate CAPR before 1983, we extrapolate it backward using the growth rates of the data by Davis et al. (2008).

[26] In the United States, mortgages represent a much larger share of debt for the non-corporate than the corporate business sector. See, e.g., the Fed's Financial Accounts of the United States at https://www.federalreserve.gov/releases/z1/.

[27] Based on the standard present-value approach to asset pricing, house prices should equal discounted expected future rents (i.e., earnings on housing assets). Accordingly, the ratio should reflect expectations on housing returns (i.e., the discount factor, given by a risk-free interest rate plus a housing risk premium) and the growth rate of rents. Like its stock-market counterpart, i.e., the price–earnings ratio, the price–rent ratio is commonly used to gauge whether assets are undervalued or overvalued, its long-term average serving as a benchmark to detect possible deviations of house prices from fundamental levels and to assess downside risks to house prices (e.g., Philiponnet and Turrini 2017).

forward-looking nature of asset pricing, the price-to-rent ratio combines information on current credit market conditions and expectations on future house prices (and thus the value of collateral). In particular, a high price–rent ratio tends to be associated with lax credit constraints and expected house price depreciation. Such a combination is likely to provide valuable predictive information on future business-cycle conditions, e.g., by signaling households' and firms' vulnerability to credit-tightening shocks and potential deleveraging. (Conversely, a low price–rent ratio may indicate that there is room for a future relaxation of credit constraints and an increase in housing wealth.) In an influential contribution, Favilukis et al. (2017) proposed a general equilibrium model that helps rationalize such dual information provided by the price–rent ratio. During economic expansions, higher house prices (collateral values) allow households to borrow more easily, thus providing greater insurance against income risk. Accordingly, the housing risk premium decreases, pushing house prices further up. At the same time, higher housing demand prompts more residential investment, which lowers the expected growth rate of rents. As a result, a higher price–rent ratio can be only justified by lower expected housing returns (discount rates), in the form of future house price depreciation, rather than faster rental growth.[28] The result that a high price–rent ratio tends to be associated with subsequent declines in house prices is also compatible with theoretical models of house price bubbles. Abraham and Hendershott (1996) propose a model in which extrapolative (i.e., backward-looking) expectations on house prices generate bubbles in good times. However, when prices become too high relative to fundamentals, expectations switch to negative, triggering a bust. Duca et al. (2011) find evidence of both time-varying credit constraints and extrapolative house price expectations as drivers of the U.S. price-to-rent ratio. The predictive power of the ratio may also be related to its direct effect on home-ownership decisions (Sommer & Sullivan, 2018): when the price–rent ratio increases, home ownership becomes more expensive relative to renting. This may lead households to prefer renting over ownership, with negative effects on housing market activities and the business cycle in general.[29]

Finally, Campbell and Shiller (1998) suggest calculating cyclically adjusted valuation ratios, i.e., dividing asset prices by multi-year averages of earnings, as a way to smooth out noise in fundamentals and thus achieve more robust valuation. This motivates the use of the CAPR ratio instead of the simple price–rent ratio.

[28] The model prediction that a high price–rent ratio forecasts lower house prices rather than high rental growth is consistent with the findings of a number of empirical papers, e.g., Campbell et al. (2009) and Kishor and Morley (2015). Other theoretical models in which the price–rent ratio mostly reflects credit market conditions include Chu (2014), Garriga et al. (2019), Greenwald and Guren (2021), and Justiniano et al. (2019).

[29] See, e.g., the FRED Blog article "Is the housing price-rent ratio a leading indicator?", available at https://fredblog.stlouisfed.org/2018/09/is-the-housing-price-rent-ratio-a-leading-indicator/.

6. Conclusions

In recent years, macroeconomic research has emphasized the role of financial conditions as key determinants of aggregate fluctuations. In particular, housing and debt cycles, building up in the background of business cycles, are now widely recognized to have profound and potentially disruptive effects on economic activity. This paper provides new empirical results on the role of specific financial-cycle indicators for predicting U.S. GDP over medium-term horizons (one to five years). Based on a wide variety of methodologies applied to a high-dimensional dataset, we found that two ratios have a particularly strong relationship with economic activity over subsequent years, both combining information on financial conditions and economic fundamentals: the CAPR ratio, a robust valuation ratio for the housing market; and the NNBLI ratio, capturing the debt burden of non-corporate (small) firms. High (low) values of these ratios predict low (high) output growth over the medium term. Compared to composite measures of financial conditions, these indicators appear to offer more stable predictive information on GDP across different business-cycle phases and different time periods. Also, their predictive ability appears to be consistent with macro-financial theories in which the interaction of housing market valuations and collateralized borrowing by firms and households represents a crucial transmission mechanism of economic fluctuations. The results of the paper show that a careful selection of financial-cycle indicators provides substantial value added to multi-year forecasts of GDP. Small models that include the best indicators are able to outperform more sophisticated, high-dimensional models. The CAPR and NNBLI ratios may thus be important tools for forecasting, economic analysis, and macro-prudential policy.

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. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2023.05.007.

References

Abraham, J. M., & Hendershott, P. H. (1996). Bubbles in metropolitan housing markets. *Journal of Housing Research*, *7*(2), 191–207.
Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *American Economic Review*, *109*(4), 1263–1289.
Adrian, T., Grinberg, F., Liang, N., Malik, S., & Yu, J. (2022). The term structure of growth-at-risk. *American Economic Journal: Macroeconomics*, *14*(3), 283–323.
Alessandri, P., & Mumtaz, H. (2017). Financial conditions and density forecasts for US output and inflation. *Review of Economic Dynamics*, *24*, 66–78.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92.
Banerjee, R., & Blickle, K. (2021). Financial frictions, real estate collateral and small firm activity in Europe. *European Economic Review*, *138*, Article 103823.
Banerjee, A., Marcellino, M., & Masten, I. (2005). Leading indicators for euro-area inflation and GDP growth. *Oxford Bulletin of Economics and Statistics*, *67*(s1), 785–813.
Bernanke, B. S., Gertler, M., & Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. In J. B. Taylor, & M. Woodford (Eds.), *Handbook of Macroeconomics*: *Vol. 1*, (pp. 1341–1393). Elsevier.
Borio, C. (2014). The financial cycle and macroeconomics: What have we learnt? *Journal of Banking & Finance*, *45*, 182–198.
Borio, C., & Lowe, P. (2002). Asset prices, financial and monetary stability: Exploring the nexus. *BIS working papers 114*, Bank for International Settlements.
Brunnermeier, M., & Krishnamurthy, A. (2020). The macroeconomics of corporate debt. *The Review of Corporate Finance Studies*, *9*(3), 656–665.
Campbell, S. D., Davis, M. A., Gallin, J., & Martin, R. F. (2009). What moves housing markets: A variance decomposition of the rent–price ratio. *Journal of Urban Economics*, *66*(2), 90–102.
Campbell, J. Y., & Shiller, R. J. (1998). Valuation ratios and the long-run stock market outlook. *The Journal of Portfolio Management*, *24*(2), 11–26.
Carriero, A., Clark, T. E., & Marcellino, M. (2015). Bayesian VARs: Specification choices and forecast accuracy. *Journal of Applied Econometrics*, *30*(1), 46–73.
Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97*(2), 465–480.
Chauvet, M., & Potter, S. (2013). Forecasting output. In G. Elliott, & A. Timmermann (Eds.), *Handbook of Economic Forecasting*: *Vol. 2*, (pp. 141–194). Elsevier.
Chu, Y. (2014). Credit constraints, inelastic supply, and the housing boom. *Review of Economic Dynamics*, *17*(1), 52–69.
Claessens, S., & Kose, M. A. (2017). Asset prices and macroeconomic outcomes: A survey. *BIS working papers 676*, Bank for International Settlements.
Claessens, S., Kose, M. A., & Terrones, M. E. (2011). Financial cycles: What? How? When? *NBER International Seminar on Macroeconomics*, *7*(1), 303–344.
Clark, T. E., & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, *105*(1), 85–110.
Clements, M. P., & Hendry, D. F. (2006). Forecasting with breaks. In G. Elliott, C. Granger, & A. Timmermann (Eds.), *Handbook of Economic Forecasting*: *Vol. 1*, (pp. 605–657). Elsevier.
Contessi, S., & Kerdnunvong, U. (2015). Asset bubbles: Detecting and measuring them are not easy tasks. *The Regional Economist*.
Coulson, N. E., & Kim, M.-S. (2000). Residential investment, non-residential investment and GDP. *Real Estate Economics*, *28*(2), 233–247.
D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, *28*(1), 82–101.
Davis, M. A., Lehnert, A., & Martin, R. F. (2008). The rent-price ratio for the aggregate stock of owner-occupied housing. *Review of Income and Wealth*, *54*(2), 279–284.
Davis, M. A., & Van Nieuwerburgh, S. (2015). Housing, finance, and the macroeconomy. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*: *Vol. 5*, (pp. 753–811). Elsevier.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.
Drehmann, M., Borio, C., & Tsatsaronis, K. (2012). Characterising the financial cycle: Don't lose sight of the medium term!. *BIS working papers 380*, Bank for International Settlements.
Duca, J. V., Muellbauer, J., & Murphy, A. (2011). House prices and credit constraints: Making sense of the US experience. *The Economic Journal*, *121*(552), 533–551.
Duca, J. V., Muellbauer, J., & Murphy, A. (2021). What drives house price cycles? International experience and policy issues. *Journal of Economic Literature*, *59*(3), 773–864.
Elliott, G., & Timmermann, A. (2016). *Economic forecasting*. Princeton University Press.
Favilukis, J., Ludvigson, S. C., & Van Nieuwerburgh, S. (2017). The macroeconomic effects of housing wealth, housing finance, and limited risk sharing in general equilibrium. *Journal of Political Economy*, *125*(1), 140–223.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2001). Coincident and leading indicators for the euro area. *The Economic Journal*, *111*(471), C62–C85.
Garriga, C., Manuelli, R., & Peralta-Alva, A. (2019). A macroeconomic model of price swings in the housing market. *American Economic Review*, *109*(6), 2036–2072.
Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225–2236.
Gertler, M., & Gilchrist, S. (2018). What happened: Financial factors in the Great Recession. *Journal of Economic Perspectives*, *32*(3), 3–30.
Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, *25*(4), 595–620.
Green, R. K. (1997). Follow the leader: How changes in residential and non-residential investment predict changes in GDP. *Real Estate Economics*, *25*(2), 253–270.
Greenwald, D. L., & Guren, A. (2021). Do credit conditions move house prices?. *NBER working papers 29391*, National Bureau of Economic Research.
Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291.
Hsu, N.-J., Hung, H.-L., & Chang, Y.-M. (2008). Subset selection for vector autoregressive processes using Lasso. *Computational Statistics & Data Analysis*, *52*(7), 3645–3657.
Jordà, Ò., Schularick, M., & Taylor, A. M. (2013). When credit bites back. *Journal of Money, Credit and Banking*, *45*(s2), 3–28.
Jordà, Ò., Schularick, M., & Taylor, A. M. (2016). The great mortgaging: Housing finance, crises and business cycles. *Economic Policy*, *31*(85), 107–152.
Jordà, Ò., Schularick, M., & Taylor, A. M. (2017). Macrofinancial history and the new business cycle facts. *NBER Macroeconomics Annual*, *31*, 213–263.
Justiniano, A., Primiceri, G. E., & Tambalotti, A. (2019). Credit supply and the housing boom. *Journal of Political Economy*, *127*(3), 1317–1350.
Kishor, N. K., & Morley, J. (2015). What factors drive the price–rent ratio for the housing market? A modified present-value analysis. *Journal of Economic Dynamics & Control*, *58*, 235–249.
Kiyotaki, N., & Moore, J. (1997). Credit cycles. *Journal of Political Economy*, *105*(2), 211–248.
Koenker, R., & Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, *94*(448), 1296–1310.
Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, *28*(2), 177–203.
Koop, G. M., & Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, *177*(2), 185–198.
Leamer, E. E. (2007). Housing IS the business cycle. *NBER working papers 13428*, National Bureau of Economic Research.
Leamer, E. E. (2015). Housing really is the business cycle: What survives the lessons of 2008–09? *Journal of Money, Credit and Banking*, *47*(S1), 43–50.
Marcellino, M., Stock, J. H., & Watson, M. W. (2003). Macroeconomic forecasting in the euro area: Country specific versus area-wide information. *European Economic Review*, *47*(1), 1–18.
McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.
McCracken, M. W., & Ng, S. (2021). FRED-QD: A quarterly database for macroeconomic research. *Federal Reserve Bank of St. Louis Review*, *103*(1), 1–44.
Mian, A., Rao, K., & Sufi, A. (2013). Household balance sheets, consumption, and the economic slump. *Quarterly Journal of Economics*, *128*(4), 1687–1726.
Mian, A., & Sufi, A. (2014). What explains the 2007–2009 drop in employment? *Econometrica*, *82*(6), 2197–2223.
Mian, A., & Sufi, A. (2015). *House of debt*. University of Chicago Press.

Mian, A., & Sufi, A. (2018). Finance and business cycles: The credit-driven household demand channel. *Journal of Economic Perspectives*, *32*(3), 31–58.
Mian, A., Sufi, A., & Verner, E. (2017). Household debt and business cycles worldwide. *Quarterly Journal of Economics*, *132*(4), 1755–1817.
Pain, N., Lewis, C., Dang, T.-T., Jin, Y., & Richardson, P. (2014). OECD forecasts during and after the financial crisis: A post mortem. *OECD Economics Department working papers 1107*, OECD Publishing.
Philiponnet, N., & Turrini, A. (2017). Assessing house price developments in the EU. *European economy - discussion paper 048*, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.
Piazzesi, M., & Schneider, M. (2016). Housing and macroeconomics. In J. B. Taylor, & H. Uhlig (Eds.), *Handbook of Macroeconomics*: *Vol. 2*, (pp. 1547–1640). Elsevier.
Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, *72*(3), 821–852.
Reichlin, L., Ricco, G., & Hasenzagl, T. (2020). Financial variables as predictors of real growth vulnerability. *Discussion papers 05/2020*, Deutsche Bundesbank.
Rossi, B. (2021). Forecasting in the presence of instabilities: How we know whether models predict well and how to improve them. *Journal of Economic Literature*, *59*(4), 1135–1190.
Shiller, R. J. (2015). *Irrational exuberance*. Princeton University Press.
Smith, J., & Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, *71*(3), 331–355.
Sommer, K., & Sullivan, P. (2018). Implications of US tax policy for house prices, rents, and homeownership. *American Economic Review*, *108*(2), 241–274.
Sommer, K., Sullivan, P., & Verbrugge, R. (2013). The equilibrium effect of fundamentals on house prices and rents. *Journal of Monetary Economics*, *60*(7), 854–870.
Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.
Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, *41*(3), 788–829.
Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.
Stock, J. H., & Watson, M. W. (2011). Dynamic factor models. In M. P. Clements, & D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting*, Oxford University Press.
Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*(1), 267–288.