Do Insurance Systemic Risk Measures Have Predictive Power?

Raja Bouzouita
University of Central Missouri

Mihaela Craioveanu
University of Central Missouri

ABSTRACT

Forecasting future economic downturns is an active area of research. This paper examines the relative predictive power of various insurance systemic risk measures. Monthly insurance equity returns from January 1, 1988, to December 31, 2023, are used to construct four systemic risk measures: MES, two specifications of ΔCoVaR, and the standard deviation of monthly insurance stock returns. The forecasting performance of these measures is tested on their ability to predict future economic activity as measured by the Chicago Fed National Activity Index (CFNAI) and the growth in the industrial production index. Using the Diebold-Mariano test and the Model Confidence Set as performance criteria, it is found that the Conditional Value at Risk measures are better predictors of an economic downturn than the other two systemic risk measures used. These measures could be used by regulators as an effective monitoring tool to avert future financial crises.

Keywords: Financial Crisis, Systemic Risk, Insurance Industry, Model Selection, Forecasting.

INTRODUCTION

Forecasting economic activity is of interest to policymakers in times of both economic stability and turmoil. The 2008-2009 financial crisis and the Covid-19 pandemic have renewed the interest in predictive measures of economic activity. The crises had devastating effects on the general economy such as a contraction in gross domestic product and high unemployment. During the financial crisis, there were also losses resulting from the failure of large financial institutions such as Lehman Brothers and the near downfall of AIG, one of the largest insurance companies. In the aftermath, extensive academic work has been dedicated to the measures and the impact of macroeconomic risks presented by financial firms.

Systemic risk is defined as the risk that a failure of one or more large financial institutions may cause other financial institutions to fail resulting in harmful macroeconomic effects. To closely monitor financial institutions, policymakers passed the Dodd–Frank 2010 Act. The new sweeping law has many provisions such as a process to systematically liquidate distressed financial institutions, and a monitoring device of financial institutions' systemic risk including insurance companies. If regulators were to rely on systemic risk as a screening tool or as an early warning system, then these measures should be tested on their predictive and forecasting power of future financial crisis and their impact on the economy. Relying on flawed measures may lead to the wrong policy recommendations and would impose an unwarranted burden on the insurance sector.

The insurance sector is a significant component of the U.S. financial system. Insurance companies are considered financial intermediaries. Premiums received are channeled back into the economy as investments. Insurance companies invest heavily in debt securities approaching sixty percent industry average in corporate and government bonds. Financial distress in the insurance sector may disrupt the bond market prompting businesses to cancel or delay capital investments which would hamper economic growth. Another important role of the insurance industry, especially the property and casualty segment, is to facilitate personal and business credit. These companies provide the coverage required by lending institutions as a condition for loan approval. Insurance companies facilitate credit not only for individuals and businesses but also for municipalities and financial institutions in the form of financial guarantees provided by bond insurers. Local and state governments spend the raised funds on infrastructure projects that benefit the local economy. However, insurance companies have expanded their non-insurance activities such as asset management. Cummins and Weiss (2014) find that the core activities (underwriting, reserving, claims settlement, and reinsurance) of property and casualty insurers do not contribute to systemic risk. However, the authors find evidence suggesting that some activities of life insurers, particularly group annuities and separate accounts, increase insurers' systemic risk. In addition, life insurance companies' liabilities have become riskier due to the reliance on shadow reinsurance companies mainly offshore to take advantage of lax tax codes. These reinsurance mechanisms do not transfer the risk as traditional reinsurance. Instead, the liabilities remain the responsibility of the holding company. Koijen and Yogo (2016) and Schwarcz (2015) also recognize shadow insurance as a potential threat to insurers and the financial system as it now makes up a large proportion of the reinsurance ceded by life insurers. For example, due to a lower degree of transparency of shadow insurers, these transactions may entail hidden risks that are, e.g., not reflected in the rating of the ceding company such as the funding structure and quality of shadow insurers. Life insurance liabilities were $8920.6 billion in 2023 (Board of Governors of the Federal Reserve System 2023) which is substantial compared to 12301.9 in savings deposits at U.S. depository institutions. (P& C liabilities 1,470 billion in 2023)

The increased interconnectedness of the financial services industry made the financial system more integrated and hence more susceptible to shocks and crises. Therefore, a crisis in the insurance sector may reverberate across the entire financial system. Considering the vital role that the insurance industry in the U.S. economy and its interconnection with the financial markets, the forecasting power of insurance systemic risks of economic downturns in explored. Evidence about which measure of insurance systemic risk can predict the economic impact will influence the priority that policymakers attach to reforming insurance sector policies.

This paper investigates the forecasting performance of insurance systemic risk measures as predictors of macroeconomic downturns. The Chicago Fed National Activity Index (CFNAI) as a measure of economic activity is used. Four systemic risk measures are calculated using daily equity returns for a sample of insurance companies from January 1, 1988, to December 31, 2023. Each measure is then used in a predictive regression model where the dependent variable is the CFNAI index. The methodology is applied to each systemic risk measure and the performance of the one-month ahead CFNAI forecast on a forecast sample is evaluated. Two different criteria are used when comparing the predictive power of the systemic risk measures: the Diebold-Mariano test (Diebold and Mariano, 1995) and the Model Confidence Set (MCS) procedure (Hansen et al., 2011). The robustness of the findings for the performance of systemic risk is tested on longer forecasting periods and an alternative measure of macroeconomic activity. The results show that the Conditional Value at Risk (CoVaR) measure of systemic risk performs best in predicting economic downturns regardless of the criteria used, the forecasting period, and the measure of economic activity.

LITERATURE REVIEW

Following the 2008-2009 financial crisis, extensive research has been developed to define systemic risk measures and quantify the contribution of individual financial institutions to the overall financial risk.

Danielsson et al. (2012) explore the performance of three systemic risk measures, Value at risk (VaR), Marginal Expected Shortfall (MES), and Conditional Value at Risk (CoVaR), to estimate each bank's contribution to systemic risk. The authors measure the performance of risk measures based on their measurement precision and ranking precision (how well the systemic risk measures rank banks by their contribution to the overall system risk using Spearman rank order correlation) using a sample of large U.S. banks. The authors find that none of the three systemic risk measures perform well. In random trials, the systemic risk measures' empirical distributions have large confidence intervals. Similarly, testing the performance of the systemic risk measures' ability to rank each bank's contribution to overall risk, the VaR, MES, and CoVaR are unreliable. Therefore, the authors conclude that the systemic risk measures have serious measure imprecision and should not be used as a regulatory tool.

Benoit et al. (2013) focus on the Marginal Expected Shortfall's (MES) ability to identify the financial institutions that are more likely to suffer the most, in terms of value, in the aftermath of a financial crisis. They calculate Spearman rank correlations between the equity losses and balance sheet indicators and MES. The authors find that balance sheet-based ratios such as tier 1 solvency ratio, size, and nonperforming loans as better predictors of bank losses because of the 2007-2008 financial crisis. In contrast, the MES is a poor predictor of such losses.

Brownlees et al. (2020) evaluate the performance of systemic risk measures using historical data on New York bank runs dating back to the pre-FIDIC period; an era marked by numerous financial crises allowing the authors to test their model across several bank panics rather than a single event like the financial crisis of 2007-2008. Like Benoit et al. (2013), the authors' goal is two-fold. First, test whether systemic risk measures, CoVaR and SRISK, correctly identify systematically important financial institutions (SIFIs). Second, examine these systemic risks' ability to forecast financial crises. The results from predictive regression analysis show that CoVaR and SRISK identify SIFIs up to six months before a financial crisis. The forecasting results only provide weak evidence that systemic risk measures predict financial crises. The authors did not conduct a statistical test to assess the forecasting power of the systemic risk measures and used a naïve approach entailing a comparison of the R^2 of the regression models with and without the systemic risk measures.

This paper is part of a growing literature that focuses on the effectiveness of systemic risk measures in predicting economic downturns. Few studies analyze the macroeconomic effects of systemic risk. Several articles argue that systemic risk measures have a limited ability to predict systemic risk. For instance, Allen et al. (2012) calculate CATFIN which is an average of three specifications of VaR for financial firms using return data covering the period from January 1973 through 2009. The comprehensive measure of systemic risk forecasted economic declines up to six months ahead. The predictive performance of CATFIN was unchanged using different proxies for economic activities. The study has two limitations. First, it investigates the impact of systemic risk on the macro economy by the financial system as a whole without regard to individual sectors. Second, it uses an average measure of systemic risk rather than comparing the forecasting performance across different measures.

Giglio et al. (2016) using comprehensive data covering the U.S. and Europe examine the forecasting ability of systemic risk measures in predicting macroeconomic declines. They measure macroeconomic shocks by the innovations in the industrial production index and the Chicago Fed National Activity Index (CFNAI). The results indicate that financial volatility, Conditional Value at Risk (CoVaR), Marginal Expected Shortfall (MES), and CATFIN of Allen, Bali, and Tang (2012), accurately forecast future macroeconomic declines. Interestingly, the financial volatility measure had the best prediction results. The authors also find strong evidence in favor of dimension reduction methods. The authors do not report results by industry within the wider financial sector. According to the discussed literature, there is no definitive consensus on the effectiveness of systemic risk measures in identifying systemically important financial institutions and in predicting future financial crises.

Two main differences distinguish this paper from the previous studies of Allen et al. (2012) and Giglio et al. (2016). First, these studies look at the entire financial system, while this paper investigates one segment of the financial system: the insurance industry. Second, in analyzing the forecasting performance of the systemic risk measures, superior forecasting performance as defined by Diebold and Mariano (1995) and Hansen et al. (2011) is used. The rest of the paper is structured as follows: Section 3 explains the methodology and the model used. Section 4 reports the estimation results. Section 5 contains the results on the ability of the systemic risk measures to predict economic downturns. Section 6 provides concluding remarks.

DATA, METHODOLOGY, AND MODEL

The purpose of this paper is to investigate the predictive power of various systemic risk measures on economic activity. the Chicago Fed National Activity Index (CFNAI) is used as a measure of economic activity. The CFNAI is a weighted average of 85 monthly economic indicators. The sample period starts on January 1, 1988, and ends on December 31, 2023. The same analysis is performed with industrial production growth (IP) as an alternative measure of economic activity.

To calculate systemic risk in the U.S. insurance industry, daily stock returns from CRSP are used covering the period from January 1, 1988, through December 30, 2023. Insurance companies with SIC codes 6311, 6321, 6324, 6331, and 6351 are screened. Life insurers, accident and health insurers with hospital and medical service plans, property and casualty insurers, and surety insurance companies are included. Only companies with a minimum number of 250 observations are included. The final sample contains 150 U.S. insurance companies.

In this paper, the following systemic risk measures are calculated: the marginal expected shortfall (MES), ΔCoVaR, and the standard deviation of the daily returns (SD). These are among the popular measures typically used in the literature.

The marginal expected shortfall (MES) was proposed by Acharya et al. (2010). MES reflects the exposure of a financial institution to the market's systemic risk and is defined as the expected return conditional on the market being at or below a certain threshold α.

$$MES_i = E[R_i | R \leq VaR(\alpha)] \quad (1)$$

where $VaR(\alpha)$ is the Value at risk. A standard risk level of α=5% in the estimation of the MES is chosen. The 5% worst days for the market returns (R) over the sample period are considered and the average return of any given firm (R_i) for these days is calculated.

Adrian and Brunnermeier (2011) propose an alternative measure of systemic risk as given by the Conditional Value-at-Risk (CoVaR). CoVaR measures the financial sector's Value at Risk (VaR) given that a particular institution is in distress.

$$Pr\ (R_{system} < CoVar_i \ | R_i = VaR_i) = \alpha \quad (2)$$

where α=5th percentile.

The ΔCoVaR for a firm *i* is defined as the difference between the VaR for the financial system conditional on this particular firm being in financial distress, and the VaR of the financial system conditional on firm *i* being in its median state.

$$\Delta CoVaR_i = CoVaR_i\ (\alpha) - CoVaR_i\ (0.5) \quad (3)$$

Following Adrian and Brunnermeier (2011), ΔCoVaR is estimated using two estimation methods: the quantile regression model ($\Delta CoVaR_{Quant}$), and the dynamic conditional correlation DCC GARCH ($\Delta CoVaR_{DCC}$) model of Engle (2002).

The three above systemic risk measures were calculated for each insurance company and then for each measure, was averaged across all the insurance companies to obtain one daily measure of systemic risk. Next, the monthly average is taken to obtain aggregate measures. The monthly conversion was done to match the frequency of the CFNAI data. The last measure of systemic risk is financial volatility (SD). For each insurance company, the within–month standard deviation of daily equity returns was calculated and then the standard deviations across the insurance companies were averaged to obtain a monthly measure.

Summary statistics for CFNAI, IP, and the systemic measures are presented in Table 1 (Appendix). The CFNAI index average is negative signifying a below-trend growth for our sample period. To understand the numbers, consider the mean of the MES. When the market returns are at their worst 5%, the average return of insurance firms is 1.86%. $CoVaR_{Quant}$ and $CoVaR_{DCC}$ have very similar averages for the sample period (0.11%). An insurance company that is in financial distress (the return of the institution is below its 5% VaR), will increase, on average, the VaR of the insurance system by 0.11% over the benchmark VaR.

Figure 1 (Appendix) shows the four monthly systemic risk measures calculated over the sample period. Consistently, the four measures of systemic risks: MES, the two CoVaR specifications, and volatility registered dramatic increases in 1992-1993, 2002-2003, and in recession periods, particularly during the 2008-09 financial crisis, the COVID-19 pandemic and the War in Ukraine.

In March 2020, the US declared COVID-19 a national emergency. Around this period, the four measures systemic risk measures soared almost reaching levels experienced during the Great Recession. The markets reacted strongly fearing an economic recession due to the ensuing lockdowns and restrictions. As the fears about recession eased thanks to the fiscal and monetary measures, volatility receded. As the number of COVID cases kept rising, there were concerns about the claims against insurance companies by businesses claiming business interruption insurance as a flurry of lawsuits were filed against property-liability insurance companies. For the most part, insurance companies fended off those claims. After the SARS (2003) pandemic insurance companies changed the wording of the business interruption insurance coverage to exclude losses emanating from viruses.

The insurance industry experienced high systemic risk levels similar to those of the financial crisis in 1992-1993. During this period, there were two major crises. The property-liability sector was hit by catastrophic losses from Hurricane Andrew in 1992. During the same time, life insurance companies suffered major losses from the collapse of the high-yield bond market (Fenn and Cole, 1994). Overall, the four measures of systemic risk follow the same pattern.

Given the notable events that took place during the sample period, the presence of structural breaks in the relationship between the measures of economic activity and the four systemic risk measures was tested. These dramatic events may have caused the data to shift. Failure to account for the changes in the data-generating process may result in inaccurate inferences and wrong policy recommendations. Bai and Perron's (1998, 2003) test was used to detect any structural breaks in the data set. The null hypothesis of no breaks was rejected. The model finds the optimum number of breaks sequentially with a series of F-test ratios that compare the SSR for the t breaks model versus the t+1 breaks model. For example, in testing for the presence of one break, the F ratio is the ratio between the SSR (Sum of Squared Regression) for 0 breaks over the SSR for one break. The test results show the presence of six structural breaks. The results for the two measures of economic activity and systemic risk measures used are presented in Tables 2 and 3 (Appendix).

Some of those breaks were matched with global events and events specific to the insurance industry. The segment from 2000-2003 can be attributed to the passage of GLBA (Gramm-Leach-Bliley Act in 1999) which allowed commercial banks, securities firms, and insurance companies to offer financial services with limited restrictions on insurance underwriting and other insurance activities. GLBA resulted in higher competition from non-insurance rivals threatening a lucrative niche market. The change between segments three and four coincides with the 2008-09 financial crisis and the change between segments five and six corresponds to the COVID-19 pandemic.

Next, the predictive power of systemic risk measures is estimated using the following model:

$$Y_t = \alpha + \beta SR_{t-1} + \sum_{i=1}^{3} \gamma_i Y_{t-i} + \varepsilon_t \tag{4}$$

where Y_t is the monthly Chicago Fed National Activity Index (CFNAI) or industrial production (IP), SR_t is the monthly systemic risk (measured in four different ways: MES, $CoVaR_{Quant}$, $CoVaR_{DCC}$, and SD), and ε_t: $WN(0,\sigma_\varepsilon^2)$.

Up to three lags for the dependent variable are allowed in the specification model. To choose the appropriate number of lags, the AIC and SIC selection criteria are used. Newey-West standard errors to account for autocorrelation and heteroscedasticity and obtain consistent estimates are employed.

It is well known in the time series literature that non-stationary variables should not be used in regressions due to the possibility of spurious results. Stationary tests for the variables considered are performed. The CFNAI is constructed as an index by using principles components with the components included being transformed to be stationary, therefore the CFNAI is stationary by construction. For the industrial production index, the growth rate which becomes stationary by construction is used. In addition, these conjectures are checked by performing the Kwiatkowski– Phillips–Schmidt–Shin (KPSS) test. The results indicate stationarity for all the variables considered for each segment.

Concerning the model specification, the objective is not to identify the factors that forecast economic downturns. The goal is to investigate the relative accuracy of systemic risk in forecast macroeconomic downturns.

ESTIMATION RESULTS

The sample period of January 1988 through December 2022 is divided into two parts: an estimation sample (January 1988 through December 2022) and a forecast sample (January 2023 through December 2023). Model (4) as described in Section III is used. Three lags of the dependent variable in the model are included in the specification as determined by the AIC and SIC model selection criteria. The model is estimated for each systemic risk measure. To take account of the structural breaks detected in Tables 2 and 3 (Appendix), dummy variables that correspond to each of the identified regimes in Equation 4 are included. The forecast performance of each specification is then evaluated on the forecast sample.

Table 4 (Appendix) presents the results for the estimation of the model for each systemic risk measure considered. Standard errors are calculated according to Newey and West (1987). Table 4 (Appendix) shows that the estimated coefficients of MES, $CoVaR_{Quant}$, $CoVaR_{DCC}$, and volatility are negative and highly significant. The R^2 is about the same for the four different measures of systemic risk.

To check the robustness of the results, the growth in industrial production seasonally adjusted (IP) is used, to remove short-term fluctuations associated with the business cycle, as an alternative measure of economic activity. The results are reported in Table 5 (Appendix).

Table 5 (Appendix) results show that the estimated coefficients for the measures of systemic risk have the predicted negative sign and are highly significant. Compared to Table 4 (Appendix) results, the results are very similar.

FORECAST PERFORMANCE

The forecast sample period (January 2023 through December 2023) is used to compare the forecasting performance of the various systemic risk measures used to predict economic downturns.

The Diebold and Mariano (1995) test of equal predictive ability to perform pairwise comparisons between the forecast performance of the systemic risk measures considered is used. Table 6 (Appendix) presents the one-month, three-month, and six-month-ahead forecast errors as well as the results of the Diebold Mariano test. Columns 2 and 5 of Panel A represent the mean-squared error (MSE) of the one-month ahead forecast for the first model as a percentage of the mean-squared error of the second model considered. A value less (greater) than one for this relative MSE indicates superior (inferior) performance of the first model compared to the second model. Columns 3 and 6 report the p-value of the Diebold-Mariano test. Under the null hypothesis, the expected loss of the two measures considered is the same. In the case of a rejection of the null hypothesis, columns 4 and 7 in Table 5 (Appendix) report the systemic risk measure that has a better forecasting performance. The loss function considered is the quadratic loss function. When comparing the MES, $\Delta CoVaR_{Quant}$, and $\Delta CoVaR_{DCC}$ against the SD, each one of the three systemic risk measures outperforms the SD at the five percent significance level. The $\Delta CoVaR_{DCC}$ outperforms the MES, and the test finds no gains in the forecasting performance between $\Delta CoVaR_{Quant}$ and the $\Delta CoVaR_{DCC}$. Finally, the test indicates no difference in the forecasting performance when comparing the MES and $\Delta CoVaR_{Quant.}$ For the IP measure of economic activity, no evidence of superior forecasting performance is found for any of the four measures considered at the five percent level of significance. However, at the ten percent level of significance, the IP results are consistent with the results for the CFNAI and the MES, $\Delta CoVaR_{Quant}$, and $\Delta CoVaR_{DCC}$ outperform the SD measure.

The robustness of the systemic risk indicators to longer forecasting periods is examined. Panels B and C of Table 6 (Appendix) report the results for three-month and six-month ahead out-of-sample forecasts, respectively. The results for the three-month ahead forecasts are in line with the results of the one-month ahead forecasts. As the forecasting period is extended, $\Delta CoVaR_{Quant}$ and $\Delta CoVaR_{DCC}$ dominate the other systemic risk measures as shown in Panel C of Table 6 (Appendix). The test results are consistent regardless of the economic activity measure used.

One of the drawbacks of the Diebold-Mariano test is that it does not allow the comparison of several models at the same time. It forces the reliance on pairwise comparison of the forecasting performance. A rejection of the null hypothesis identifies the model that has a better forecasting performance, but it is hard to find a set of models that are considered to have the best performance. This problem is solved by looking at tests that allow the comparison of multiple forecasts simultaneously. To test if there are one or several measures with a significant forecasting performance compared to the other ones, the Model Confidence Set (MCS) procedure of Hansen et al. (2011) is employed. The MCS starts with a set of models that includes all the competing models (the model with $\Delta CoVaR_{DCC}$, $\Delta \mathrm{CoVaR}_{Quant}$, MES, and SD as explanatory variables). The null hypothesis is that all models have equal predictive ability. The alternative states that one of the models has a forecasting performance that is not as good as the forecasting performance of the other models considered, as quantified by a quadratic loss function.

Let M_0 be a set containing m_0 competing forecasting models. The relative performance of any two models i and j at time t is quantified by the loss differential:

$$d_{ij,t} = L_{i,t} - L_{j,t} \text{ with } i, i \in M_0, t = 1, \dots, n \tag{5}$$

The set of superior models is given by:

$$M^* = \{i \in M_0 : E(d_{ij,t}) \leq 0 \; for \; j \in M_0\} \tag{6}$$

For a given subset $M \subset M_0$, the Null and Alternative hypotheses are given as:

$$\begin{matrix} H_{0,M}: E(d_{ij,t}) = 0 \\ H_{1,M}: E(d_{ij,t}) \neq 0 \end{matrix} \quad \text{with } i, j \in M \tag{7}$$

The test statistics for the loss differential between the models is given by:

$$T_R = max_{i,j \in M} |t_{ij}| \tag{8}$$

with

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{Var(\bar{d}_{ij})}} \tag{9}$$

and the sample loss difference between models i and j is defined as:

$$\bar{d}_{ij} = \frac{1}{n} \sum_{t=1}^{n} d_{ij,t} \tag{10}$$

In the case the null is rejected the procedure eliminates the model with the worst forecasting performance. The elimination rule for the test statistic in (8) is given by

$$e_{R,M} = argmax_{i \in M} sup_{j \in M} t_{ij} \tag{11}$$

The test is performed iteratively and on each iteration the model with the worst forecasting performance is eliminated until there are no remaining models to be eliminated and cannot reject the null for the set of models belonging to the set of superior models M^*. The final set of surviving models consists of the set of superior or best models for a given level of confidence.

A confidence level of 95% is used. Table 7 (Appendix) presents the p-values associated with the MCS for our four measures of systemic risk. A model with a p-value greater or equal to the significance level α will be included in the final set of superior models. Using the one-month ahead forecast, the null hypothesis is rejected for the MES, $\Delta CoVaR_{Quant}$, and SD. The model with $\Delta CoVaR_{DCC}$ turns out to be the best MCS model for both the CFNAI and the IP indexes. These results are similar to the results from the Diebold-Mariano test which identified the $\Delta CoVaR_{DCC}$ as having a superior performance when compared with the MES and SD but was not able to distinguish between the forecasting performance of the $\Delta CoVaR_{Quant}$ and the forecasting performance of $\Delta CoVaR_{DCC}$.

Panel B and panel C of Table 7 report the results of MCS using three-month and six-month out-of-sample forecasts. The three-month forecasts are similar to the one-month forecast confirming the superiority of $\Delta CoVaR_{Quant}$ and $\Delta CoVaR_{DCC}$ as predictors of economic downturns triggered by insurance systemic risk using industrial production as the measure of economic activity with $\Delta CoVaR_{Quant}$ ranked first. For the six-month forecasts only the $\Delta CoVaR_{DCC}$ is included in the MCS. The IP selects all three measures of systemic risk (MES, $\Delta CoVaR_{DCC}$, $\Delta CoVaR_{Quant}$). In conclusion, the results from MCS further corroborate the superior predictability of insurance returns' volatility.

Overall, the results are different from the findings of Giglio et al. (2016) who found that financial volatility was the best predictor of macroeconomic downturns compared to other more sophisticated measures (nineteen in all) of systemic risk. For our sample period, the $\Delta CoVaR_{DCC}$ systemic measure is the best predictor when compared to the other measures considered (MES, SD, $\Delta CoVaR_{Quant}$).

SUMMARY AND CONCLUSION

Recent episodes of economic instability have renewed interest in predictive measures of economic activity. Both the financial crisis and the pandemic have caused financial instability in the economy. Academicians and policymakers embarked on a mission to understand systemic risk as a way to prevent future crises or limit their impact. The insurance industry represents a sizable share of the economy, as financial intermediaries, credit facilitators, and financiers of potentially catastrophic losses.

This paper investigates whether four popular insurance systemic risk measures forecast macroeconomic downturns. Two different approaches are used to check the robustness of the results and search for the best forecasting performance among the four systemic risk measures. The Diebold Mariano (1995) test is performed that allows pairwise comparisons and the Model Confidence Set (Hansen et al. 2011) which allows for multiple comparisons at the same time. It is found that the measures of $\Delta \text{CoVaR}_{DCC}$ and $\Delta CoVaR_{Quant}$ have the best forecasting performance when compared to the other measures considered. The results hold for three-month and six-month forecasting horizons and are invariant to the measure of economic activity.

The findings have important implications. Policymakers, regulators, and insurance professionals (executives) can monitor insurance systemic risk. The ability to assess insurance companies' systemic risk and its economic impact gives the regulatory authorities the capacity to implement effective policies that would reduce the exposure of the economy to systemic risk emanating from the insurance sector. Also, if a financial crisis takes place, those policymakers can help stabilize financial markets and reduce the macroeconomic impact, as these financial crises require significant resources to fix the ensuing damage. The measure of systemic risk that was found as a robust predictor of economic downturns, among other tools, can be used as a warning device to avert a financial crisis.

The findings provide empirical support for the Financial Services Oversight Council established by the Dodd-Frank Act 2010 to identify and regulate systemically important insurance companies and/or groups of insurers. This research finds evidence that there is a linkage between financial stability in the insurance sector and the real economy. The scope of regulation should expand to the entire insurance sector as there is potential for correlations among individual insurance companies that could cause or contribute to a widespread financial crisis. The lessons learned from the financial crisis experience will contribute to the development of best practices, and regulatory frameworks. The current state regulation may not be adequate. For instance, the individual states do not have jurisdictions across state lines. Structurally, insurance companies are organized as single entities or members of an insurance group. State regulation focuses on individual entities of a group and independent single insurers rather than insurance groups. In other words, state insurance regulation does not address the operations of insurance holding companies.

REFERENCES

Allen, L., Bali, T. G., & Tang, Y. (2012). Does systemic risk in the financial sector predict future economic downturns?. *The Review of Financial Studies*, *25*(10), 3000-3036.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, *9*(3), 203-228.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47-78.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, *18*(1), 1-22.

Bemanke, B., & Gertler, M. (1989). Agency costs, net worth, and business fluctuations. *American Economic Review*, *79*(1), 14-31.

Bernanke, B. S., Gertler, M., & Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics*, *1*, 1341-1393.

Bisias, D., Flood, M., Lo, A. W., & Valavanis, S. (2012). A survey of systemic risk analytics. *Annual Review of Financial Economics*, *4*(1), 255-296..

Bouzouita, R., Craioveanu, M., & Young, A. (2018). Systemic risk in the insurance industry and its impact on the economy. *Journal of Finance Issues*, *17*(1), 40-48.

Benoit, S., Colletaz, G., Hurlin, C., & Pérignon, C. (2013). A theoretical and empirical comparison of systemic risk measures. *HEC Paris Research Paper No. FIN-2014-1030*.

Brownlees, C., Chabot, B., Ghysels, E., & Kurz, C. (2020). Back to the future: Backtesting systemic risk measures during historical bank runs and the great depression. *Journal of Banking & Finance*, *113*, 105736.

Cummins, J. D., & Weiss, M. A. (2014). Systemic risk and the US insurance sector. *Journal of Risk and insurance*, *81*(3), 489-528.

Danielsson, J., James, K. R., Valenzuela, M., & Zer, I. (2012). Dealing with systemic risk when we measure it badly. *European Center for Advanced Research in Economics and Statistics Working Paper*.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), 134-144.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339-350.

Fenn, G. W., & Cole, R. A. (1994). Announcements of asset-quality problems and contagion effects in the life insurance industry. *Journal of Financial Economics*, *35*(2), 181-198.

Fields, J. A., Ross, J. B., Ghosh, C., & Johnson, K. B. (1994). Junk bonds, life insurer insolvency, and stock market reactions: The case of first executive corporation. *Journal of Financial Services Research*, *8*, 95-111.

Giglio, S., Kelly, B., & Pruitt, S. (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics*, *119*(3), 457-471.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453-497.

He, Z., & Krishnamurthy, A. (2014). *A macroeconomic framework for quantifying systemic risk* (No. w19885). National Bureau of Economic Research.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. *Journal of Econometrics*, *54*(1-3), 159-178.

Newey, W., and K. West (1987). A Simple, Positive-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55, no. 3 : 703-708.

Scaillet, O. (2005). Nonparametric Estimation of Conditional Expected Shortfall. *Insurance and Risk Management Journal* 74, no. 1: 639-660.

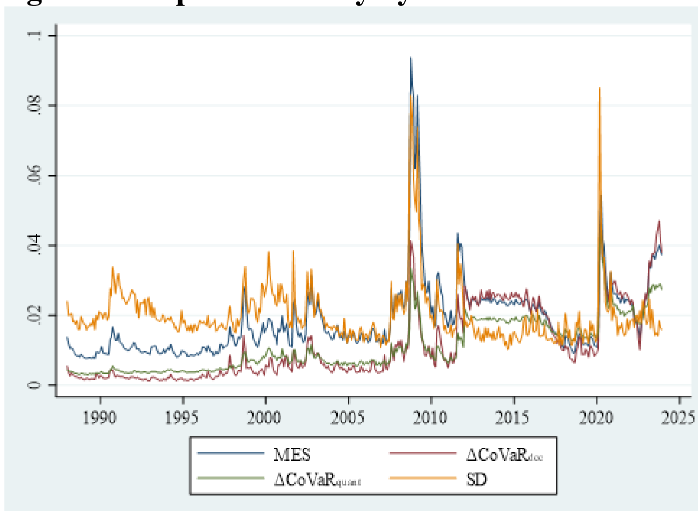
Schwarcz, D. (2015). The risks of shadow insurance. *Ga. L. Rev.*, *50*, 163.

Koijen, R. S., & Yogo, M. (2016). Shadow insurance. *Econometrica*, *84*(3), 1265-1287.

Ward, D., & Zurbruegg, R. (2000). Does insurance promote economic growth? Evidence from OECD countries. *Journal of Risk and insurance*, 489-506.

Appendix

Figure 1 Graph of Monthly Systemic Risk Measures



The figure plots the systemic risk measures from January 1988 until December 2023.

Table 1 Summary Statistics for the sample period January 1988 to December 2023

	Economic Measures		Systemic Risk Measures			
	IP	**CFNAI**	**MES**	**$\Delta CoVaR_{Quant}$**	**$\Delta CoVaR_{DCC}$**	**SD**
N	432	432	432	432	432	432
Mean	0.0013	-0.0658	0.0186	0.0110	0.0113	0.0201
STD	0.0100	1.1029	0.0112	0.0075	0.1039	0.0087
Skewness	-5.4830	-9.7179	2.9336	1.0832	1.1173	3.8362
Kurtosis	83.2870	169.4881	16.0490	3.3481	3.3481	23.8522

The table reports summary statistics for the economic and systemic risk measures.

Table 2 Bai and Perron (2003) Test for Structural Breaks, CFNAI				
SR	MES	$\Delta CoVaR_{DCC}$	$\Delta CoVaR_{Quant}$	SD
Panel A				
UDmax	113.44***	62.17***	183.71***	258.91***
F(1|0)	29.98***	24.15***	48.86***	55.34***
F(2|1)	281.8***	172.26***	325.13***	372.51***
F(3|2)	177.74***	210.43***	84.78***	32.77***
F(4|3)	186.01***	247.32***	99.55***	39.90***
F(5|4)	193.85***	255.37***	107.10***	41.89***
F(6|5)	197.96***	259.75***	110.54***	43.14***
Panel B				
Segment 1				
Start Break	1/1/1988	1/1/1988	1/1/1988	1/1/1988
End Break	1/1/1992	1/1/1992	1/1/1992	1/1/1992
Segment 2				
Start Break	1/2/1992	2/1/1992	2/1/1992	2/1/1992
End Break	12/1/1999	12/1/1999	6/1/1996	3/1/2000
Segment 3				
Start Break	1/1/2000	1/1/2000	7/1/1996	4/1/2000
End Break	7/1/2003	7/1/2003	3/1/2000	10/1/2005
Segment 4				
Start Break	8/1/2003	8/1/2003	4/1/2000	11/1/2005
End Break	2/1/2009	5/1/2009	5/1/2009	5/1/2009
Segment 5				
Start Break	3/1/2009	6/1/2009	6/1/2009	6/1/2009
End Break	8/1/2016	8/1/2016	8/1/2016	8/1/2016
Segment 6				
Start Break	9/1/2016	9/1/2016	9/1/2016	9/1/2016
End Break	3/1/2020	3/1/2020	3/1/2020	3/1/2020
Segment 7				
Start Break	4/1/2020	4/1/2020	4/1/2020	4/1/2020
End Break	12/1/2023	12/1/2023	12/1/2023	12/1/2023

The Table reports in panel A the UDmax statistic for the test of the null hypothesis of zero breaks against the alternative hypothesis of an unknown number of breaks and the sequential F ratios tests. Panel B shows the start date and the end date of the different regimes detected.

Table 3 Bai and Perron (2003) Test for Structural Breaks, IP Growth

SR	MES	$\Delta CoVaR_{DCC}$	$\Delta CoVaR_{Quant}$	SD
Panel A				
UDmax	56.11***	36.19***	79.71***	88.59***
F(1|0)	17.71***	17.17***	34.02***	29.95***
F(2|1)	120.57***	84.73***	125.06***	129.24***
F(3|2)	65.67***	87.11***	34.41***	16.50***
F(4|3)	67.85***	94.52***	37.06***	16.19***
F(5|4)	68.76***	95.83***	38..15***	16.70***
F(6|5)	69.76***	96.98***	38.65***	17.03***
Panel B				
Segment 1				
Start Break	1/1/1988	1/1/1988	1/1/1988	1/1/1988
End Break	1/1/1992	1/1/1992	1/1/1992	1/1/1992
Segment 2				
Start Break	2/1/1992	2/1/1992	2/1/1992	2/1/1992
End Break	4/1/2000	6/1/1996	9/1/1996	5/1/2000
Segment 3				
Start Break	5/1/2000	6/1/1996	10/1/1996	6/1/2000
End Break	8/1/2008	4/1/2005	4/1/2005	5/1/2009
Segment 4				
Start Break	9/1/2008	5/1/2005	5/1/2005	6/1/2009
End Break	3/1/2012	5/1/2009	4/1/2009	1/1/2013
Segment 5				
Start Break	4/1/2012	6/1/2009	5/1/2009	2/1/2013
End Break	8/1/2016	8/1/2016	8/1/2016	8/1/2016
Segment 6				
Start Break	9/1/2016	9/1/2016	9/1/2016	9/1/2016
End Break	3/1/2020	3/1/2020	3/1/2020	3/1/2020
Segment 7				
Start Break	4/1/2020	4/1/2020	4/1/2020	4/1/2020
End Break	12/1/2023	12/1/2023	12/1/2023	12/1/2023

The Table reports in panel A the UDmax statistic for the test of the null hypothesis of zero breaks against the alternative hypothesis of an unknown number of breaks and the sequential F ratios tests. Panel B shows the start date and the end date of the different regimes detected.

Table 4 Estimation Results for CFNAI

	MES		$CoVaR_{Quant}$		$CoVaR_{DCC}$		St. Deviation (SD)	
	Coeff.	T-stat	Coeff.	T-stat	Coeff.	T-stat	Coeff.	T-stat
β	-32.6185	-5.52^{***}	-75.6713	-5.61^{***}	-49.5339	-5.21^{***}	-53.9370	-9.26^{***}
R^2	15.94%		16.34%		15.38%		26.22%	

The table reports the parameter estimates of systemic risk measures, t-statistics (based on Newey West's (1987) standard errors), and the R^2. **5% significant, ***1% significant.

Table 5 Estimation Results for IP Growth

	MES		$CoVaR_{Quant}$		$CoVaR_{DCC}$		St. Deviation (SD)	
	Coeff.	T-stat	Coeff.	T-stat	Coeff.	T-stat	Coeff.	T-stat
β	-0.1919	-4.17^{***}	-0.9849	-7.60^{***}	-0.4478	-5.20^{***}	-0.4249	-7.78^{***}
R^2	16.99%		24.73%		16.75%		23.01%	

The table reports the parameter estimates of systemic risk measures, t-statistics (based on Newey West's (1987) standard errors), and the R^2. **5% significant, ***1% significant.

Table 6 Diebold Mariano Predictive Ability Test

Panel A: One-Month	CFNAI			IP		
	MSE	p-value	Superior Measure	MSE	p-value	Superior Measure
MES vs $\Delta CoVaR_{DCC}$	1.5377	0.0377	$\Delta CoVaR_{DCC}$	1.4587	0.2678	
MES vs $\Delta CoVaR_{Quant}$	0.9694	0.7929		1.2277	0.3044	
$\Delta CoVaR_{DCC}$ vs $\Delta CoVaR_{Quant}$	0.6304	0.5558		0.8416	0.5558	
MES vs SD	0.3289	0.0007	MES	0.6050	0.0587	MES
$\Delta CoVaR_{Quant}$ vs SD	0.3393	0.0000	$\Delta CoVaR_{Quant}$	0.4928	0.0924	$\Delta CoVaR_{Quant}$
$\Delta CoVaR_{DCC}$ vs SD	0.2139	0.0000	$\Delta CoVaR_{DCC}$	0.4147	0.0842	$\Delta CoVaR_{DCC}$
Panel B: Three-Month	CFNAI			IP		
	MSE	p-value	Superior Measure	MSE	p-value	Superior Measure
MES vs $\Delta CoVaR_{DCC}$	0.2041	0.0098	$\Delta CoVaR_{DCC}$	0.8404	0.7203	
MES vs $\Delta CoVaR_{Quant}$	2.0973	0.0002	$\Delta CoVaR_{Quant}$	0.9508	0.7809	
$\Delta CoVaR_{DCC}$ vs $\Delta CoVaR_{Quant}$	1.0985	0.7828		1.1314	0.7519	
MES vs SD	0.3128	0.0006	MES	0.4190	0.0242	MES
$\Delta CoVaR_{Quant}$ vs SD	0.1492	0.0001	$\Delta CoVaR_{Quant}$	0.4407	0.0770	$\Delta CoVaR_{Quant}$
$\Delta CoVaR_{DCC}$ vs SD	0.1638	0.0001	$\Delta CoVaR_{DCC}$	0.4986	0.2803	
Panel C: Six-Month	CFNAI			IP		
	MSE	p-value	Superior Measure	MSE	p-value	Superior Measure
MES vs $\Delta CoVaR_{DCC}$	3.1075	0.0429	$\Delta CoVaR_{DCC}$	0.9097	0.8691	
MES vs $\Delta CoVaR_{Quant}$	1.7554	0.0003	$\Delta CoVaR_{Quant}$	1.0498	0.6446	
$\Delta CoVaR_{DCC}$ vs $\Delta CoVaR_{Quant}$	0.5649	0.3298		1.1540	0.7837	
MES vs SD	0.2284	0.0003	MES	0.3926	0.0290	MES
$\Delta CoVaR_{Quant}$ vs SD	0.1301	0.0001	$\Delta CoVaR_{Quant}$	0.3740	0.0416	$\Delta CoVaR_{Quant}$
$\Delta CoVaR_{DCC}$ vs SD	0.0735	0.0001	$\Delta CoVaR_{DCC}$	0.4316	0.2556	

Null Hypothesis: the forecasting performance of the two models is the same. For each measure of economic activity considered (CFNAI and IP), the second and fifth columns report the mean-squared-error (MSE) of the month(s) ahead forecast, the third and sixth columns report the Diebold-Mariano p-value, and the fourth and seventh column report the chosen measure (according to the MSE criterion) in case of rejection of the null hypothesis.

Table 7 Hansen Model Confidence Set

Superior Set of Models **Panel A**: One-Month Forecasts	CFNAI		IP	
	MSE	p-value	MSE	P-value
$\Delta CoVaR_{DCC}$	0.2473	1.0000*	0.0000	1.0000*
MES	0.3802	0.0000	0.0000	0.0130
$\Delta CoVaR_{Quant}$	0.3922	0.0000	0.0000	0.1310*
SD	1.1560	0.0000	0.0000	0.0120
Panel B: Three-Month Forecasts	CFNAI		IP	
	MSE	p-value	MSE	P-value
$\Delta CoVaR_{DCC}$	0.0416	0.3460*	0.0000	0.3640*
MES	0.1518	0.0000	0.0000	0.0000
$\Delta CoVaR_{Quant}$	0.0345	1.0000*	0.0000	1.0000*
SD	1.5522	0.0000	0.0000	0.0000
Panel C: Six-Month Forecasts	CFNAI		IP	
	MSE	p-value	MSE	P-value
$\Delta CoVaR_{DCC}$	0.0095	1.0000*	0.0000	0.2790*
MES	0.0917	0.0000	0.0000	0.2790*
$\Delta CoVaR_{Quant}$	0.0297	0.0450	0.0000	1.0000*
SD	1.7585	0.0000	0.0000	0.0000

For each measure of economic activity considered (CFNAI and IP), the second and fourth columns report the mean-squared-error (MSE) of the month(s) ahead forecast, and the third and fifth columns report the MCS p-values for the four systemic risk measures considered. The Null Hypothesis states that there is no difference in predictive ability. *represents inclusion in the 95% MCS.

8%G