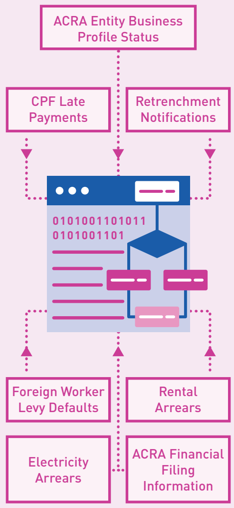
MONITORING CORPORATE HEALTH IN SINGAPORE USING A MACHINE LEARNING MODELLING APPROACH

OVERVIEW

As part of macroeconomic surveillance, it is important to monitor the corporate health of firms in the economy on a timely basis. This feature article (i) describes the development of a Corporate Health Index (CHI) to monitor the health of firms in Singapore by applying predictive machine-learning approaches to high-frequency firm-level data; and (ii) examines recent trends of the corporate health for the overall economy and for selected sectors.

METHODOLOGY

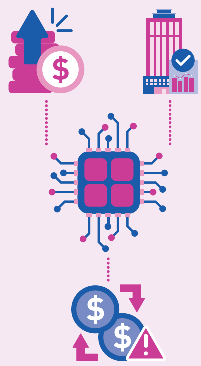
Our study draws on datasets from five government agencies to construct our CHI model. Our main target variable is the proportion of firms in distress, which we use as a measure of corporate health for a particular sector or type of firms. In addition, we make use of several high-frequency, firm-level indicators that capture information on firms' cashflow situation as input variables. These indicators are: (i) CPF Late Payments, (ii) Retrenchment Notifications, (iii) FWL Defaults, (iv) JTC Rental Arrears and (v) Electricity Arrears. Another data series that we make use of as an input variable for our model is the annual financial filing information from ACRA records, which includes the firm's revenue, total assets, share capital, and various cashflows.



We bring together these various data sources and train our CHI model to extract the relationship between firms' entity status and their various indicators of financial health and abilities to meet their cashflow demands. Using our CHI model, we can calculate a distress score for each firm, which is the predicted probability of each firm in our dataset to be in distress within a three-month period. We then aggregate these individual firm-level distress scores, by taking a weighted mean (weighted by the employment size of each firm) across firms to derive the CHI in a particular sector.

FINDINGS

In terms of variable importance, total non-current assets, net profit after taxes, and total liabilities have the highest variable importance scores — this indicates that they carry the highest informational value for our ensemble model when they predict for the target variable. Among the high-frequency firm-level indicators, CPF Late Payments and Retrenchment Notifications have comparatively much higher variable importance scores than FWL Defaults, JTC Rental Arrears and Electricity Arrears.



Using our CHI model, we find that the overall CHI for June 2023 remains stable, suggesting corporate health is likely to remain stable in the near term. However, for firms in the food & beverage services, retail trade and wholesale trade sectors, corporate health could weaken slightly, reflecting the effect of higher business costs and weaker global demand.

EXECUTIVE SUMMARY

- In this study, we develop a Corporate Health Index (CHI) to monitor corporate health in Singapore on a more timely and forward-looking basis by applying predictive machine learning approaches on a range of firm-level indicators, including higher-frequency indicators that capture information on firms' ability to meet their cashflow obligations.

- We find that corporate health in Singapore is affected by macroeconomic developments, such as the tightening and loosening of pandemic-era restrictions, as well as global interest rate hikes. We also find that the corporate health of small- and medium-sized enterprises (SMEs) is systematically weaker than that of non-SMEs.

- Looking ahead, the CHI suggests that overall corporate health in Singapore is likely to remain stable in the near term. We will continue to monitor the CHI closely in the coming months given the prevailing economic headwinds.

INTRODUCTION

As part of macroeconomic surveillance, it is important to monitor corporate health[1] on a timely basis, as firms are the main drivers of output and employment in an economy. In Singapore, indicators such as non-performing loan ratios and firm cessation are regularly used to monitor corporate health. However, these indicators either come with a lag or are not granular enough to enable the economic surveillance of specific sectors or firm types.

To overcome these limitations, we develop a Corporate Health Index (CHI) to monitor corporate health in Singapore on a more timely and forward-looking basis by applying predictive machine learning approaches on firm-level indicators, including higher-frequency indicators that capture information on firms' ability to meet their cashflow obligations. We use the CHI to provide an assessment of corporate health at the overall economy and sectoral levels, as well as for specific firm types.

LITERATURE REVIEW

The surveillance of corporate health has traditionally relied on financial ratios or estimates from regression analysis. For example, Altman (1968)'s Z-score is widely regarded as an industry benchmark for predicting corporate financial distress (Liang et al., 2020). In recent years, many academics have found that predictions of firms in distress can be significantly improved with models that use more sophisticated machine learning techniques such as support vector machines (Barboza et al., 2017; Tsai et al., 2014), tree-based algorithms (Barboza et al., 2017; Olson et al., 2012; Sun et al., 2017) and neural networks (Tsai et al., 2014). Furthermore, ensemble methods, which combine the predictions of multiple models, have been shown to outperform single model classification techniques (Graczyk et al., 2010; Liang et al., 2018; Tsai, 2014).

DATA

Our CHI model uses indicators from firms' financial statements filed annually with the Accounting and Corporate Regulatory Authority (ACRA) and monthly data from various administrative sources that reflect firms' ability to meet their cashflow obligations [Exhibit 1].

Our main variable of interest is the proportion of firms in distress, which we use as a measure of corporate health for the overall economy or a particular sector or firm type. To construct the target variable (i.e., firms in distress), we tap on ACRA's data on firms' entity status which captures the official operational status of firms across time. We identify a firm as being in distress if it takes on an entity status that indicates that the firm may be facing operational

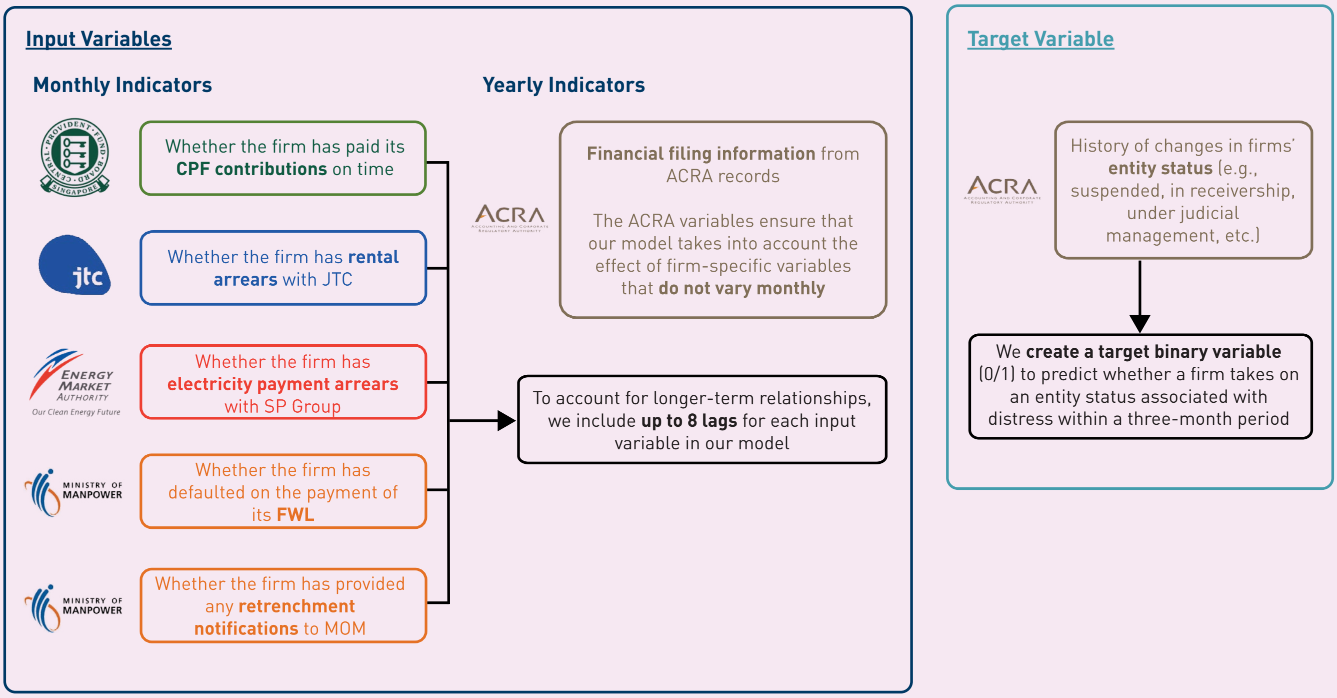
1 *In this article, corporate health refers to the ability of firms to meet their cashflow obligations and remain operational. If firms show signs of cashflow strains, they are more likely to have poorer corporate health.*

troubles or nearing cessation (e.g., "suspended", "in receivership" or "under judicial management").[2] Our objective is to predict for each firm in our dataset whether the firm will take on an entity status associated with distress within a three-month period. With these predictions for all the firms in our dataset, we can calculate the proportion of firms likely to be in distress, thereby providing a forward-looking measure of corporate health in the economy at various levels of aggregation (e.g., overall economy, sectors or firm types).

For the input variables into the model, we make use of two sets of data. The first set comprises high-frequency, firm-level indicators that capture information on firms' cashflow situation. These indicators are: (i) "CPF Late Payments" which shows whether a firm has paid its Central Provident Fund (CPF) contributions for its employees on time; (ii) "Retrenchment Notifications" which shows whether a firm has provided any retrenchment notifications to the Ministry of Manpower (MOM); (iii) "FWL Defaults" which shows whether a firm has defaulted on the payment of its Foreign Worker Levy (FWL); (iv) "JTC Rental Arrears" which shows whether a firm has rental arrears with the JTC Corporation; and (v) "Electricity Arrears" which shows whether a firm has electricity payment arrears with the Singapore Power (SP) Group.

The second set of input variables comprises the annual financial information that firms file with ACRA, including the firm's revenue, total assets, share capital and various cashflow ratios. These variables are incorporated into our model to account for firm-specific information which changes less frequently but could still affect the probability of firms falling into distress.

Exhibit 1: Types of Data Used in the CHI Model



MODEL DEVELOPMENT

We develop and deploy our CHI model using a four-stage process [Exhibit 2]. First, we assemble the input variables and carry out two standard pre-processing steps before ingesting them into our model. The first pre-processing step is to impute missing "0"s in our input data, as the coverage of firms is incomplete across the different datasets. For example, a firm could show up in the CPF Late Payments data, but not in the JTC Rental Arrears data for that month. To circumvent this issue, we impute missing variables with a "0" for numeric variables, and a "missing" category for categorical variables. For each input variable, we also generate a corresponding "missingness" variable to indicate whether the input variable was missing in our data before imputation. This first pre-processing step is important as it maximises the universe of firms and information from which our model can learn about the relationship between the input variables and the target variable.

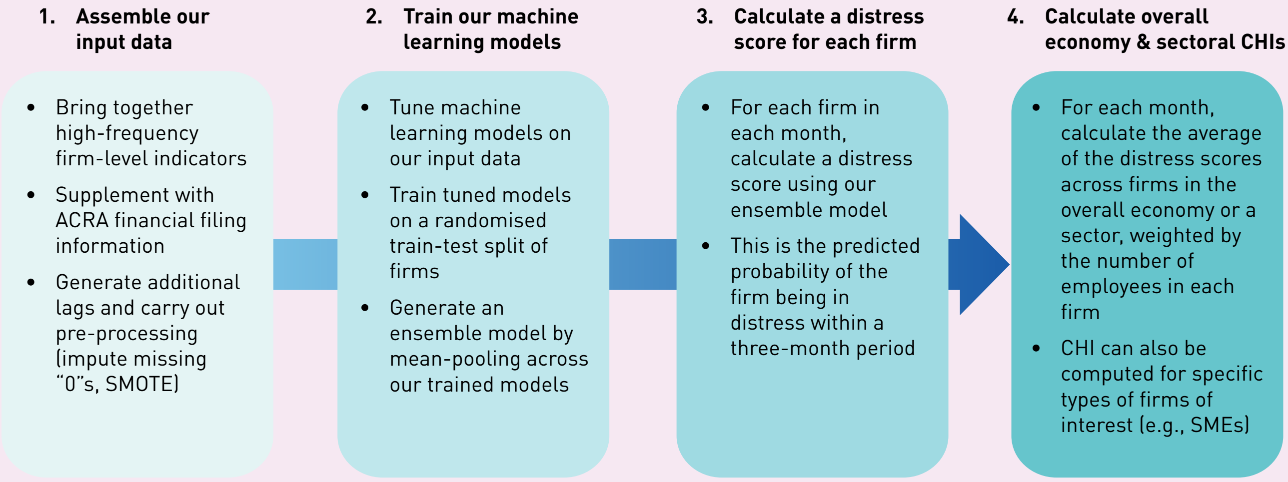
2 *The list of entity statuses in our dataset is as follows: "Ceased registration", "Cancelled", "Terminated", "Struck off", "Dissolved", "Gazetted", "Suspended", "Under judicial management", "In liquidation", "In receivership", "Amalgamated" and "Live". All these statuses, except for "Amalgamated" and "Live", are taken as statuses of distress in our model.*

The second pre-processing step is to carry out SMOTE, or Synthetic Minority Oversampling Technique, on the target variable (i.e., firms in distress).[3] As firms in distress typically form the minority of all firms in an economy, our model may not have sufficient information to learn about the relationship between the input variables and the target variable. SMOTE rectifies this issue by generating additional synthetic data points on firms in distress, thereby increasing the pool of firms in distress from which the model can learn. This in turn improves the ability of the model to predict the occurrence of such firms.

Second, with the pre-processed data, we train four machine learning models (i.e., Random Forest, XGBoost, Adaptive Boosting, Light Gradient Boosting) to extract the relationship between the input variables and the target variable. Such machine learning models tend to perform better than traditional models[4] as they do not assume any restrictions in the functional form of the relationship between the input and target variables, and are thus better able to capture non-linearities in the relationship. We also train an ensemble model which mean-pools[5] across the four models. This ensemble model constitutes our CHI model.

Third, using our CHI model, we calculate a distress score for each firm in our dataset. The distress score is the predicted probability of the firm being in distress within a three-month period. We then aggregate these individual firm-level distress scores, by taking a weighted mean (weighted by the employment size of each firm) across firms to derive the proportion of firms in distress in the overall economy or for a particular sector. Besides a sectoral perspective, we can also compute the CHI for specific types of firms, such as small- and medium-sized enterprises (SMEs).

Exhibit 2: CHI Model Development and Deployment

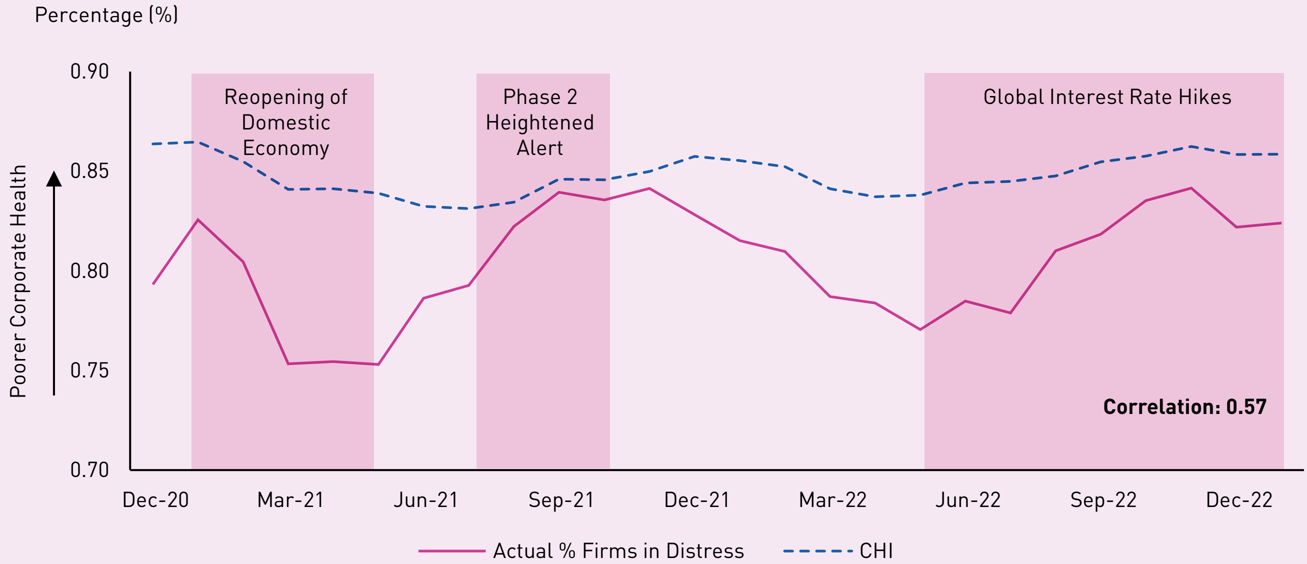


MODEL RESULTS

The results of our CHI model correlate well with the actual percentage of firms in distress based on their entity status in ACRA's data computed within a three-month forward-rolling window[6] [Exhibit 3]. In particular, the Pearson correlation coefficient between the CHI and actual percentage of firms in distress within a three-month forward-rolling window is moderately high at 0.57, indicating that our model is able to capture the broad trends of distressed firms over time. Importantly too, the movements in the CHI over time match our intuition as to how macroeconomic developments would affect the health of firms in the Singapore economy (e.g., the CHI rose when pandemic-related restrictions were imposed and fell when the restrictions were relaxed).

3 *For technical details on the SMOTE methodology, refer to Chawla et al. (2002).*
4 *Examples of traditional models are Linear Probability Models, and Logistic and Probit Regressions.*
5 *We use the F1 scores of the four individual constituent models as weights when mean-pooling for the ensemble model. The F1 score is a standard measure of a model's performance in machine learning.*
6 *As the CHI predicts the proportion of firms in distress within a three-month period, using a similar three-month forward-rolling window to compute the actual proportion of firms in distress (using ACRA data) will allow the latter to 'match' the time period captured by the CHI. For example, doing so allows us to compare the CHI reading for January 2023 (i.e., the predicted proportion of firms in distress from January 2023 to March 2023) with the actual proportion of firms in distress from January 2023 to March 2023 (i.e., a three-month forward-rolling window from January 2023).*

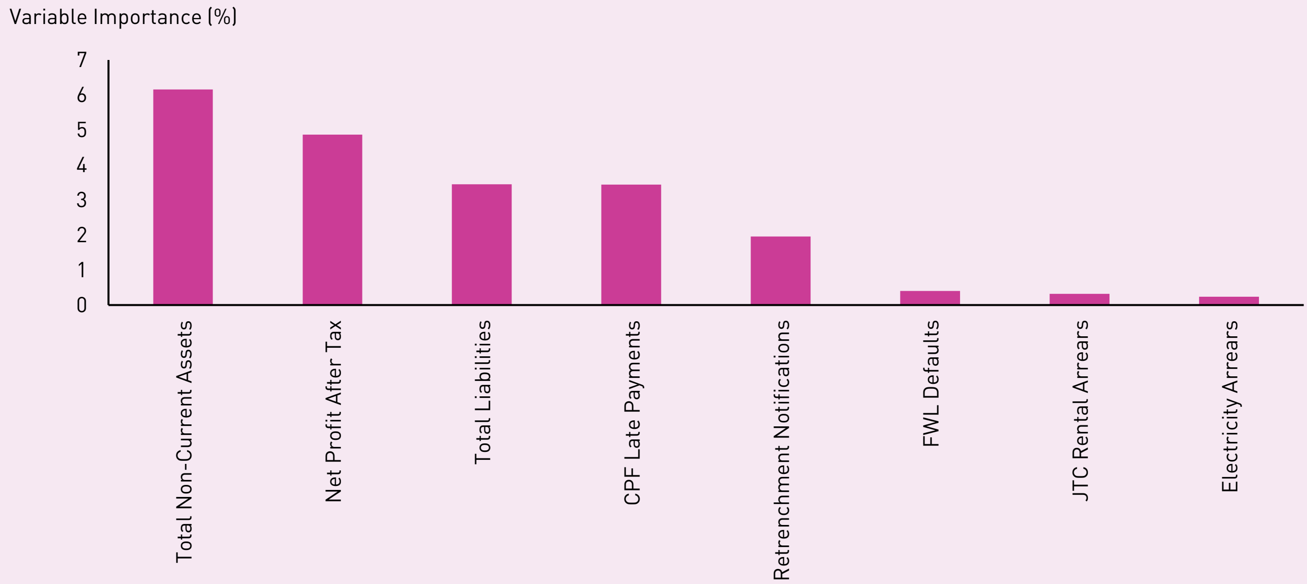
Exhibit 3: Comparison of Actual Percentage of Firms in Distress* and CHI^, Overall Economy



Notes: * The actual percentage of firms in distress is defined as the share of firms with an entity status that is associated with distress (see footnote 2) in ACRA's data, and is calculated based on a three-month forward-rolling window. For example, the data point for January 2023 shows the actual percentage of firms in distress in the months of January, February and March 2023.
^ The CHI shown for each month represents the predicted probability of firms being in distress within a three-month period. For example, the CHI for January 2023 represents the predicted probability of firms being in distress in the months of January, February and March 2023.

In terms of variable importance[7], total non-current assets, net profit after taxes and total liabilities have the highest variable importance scores. In other words, these variables have been found to carry the most informational value for our ensemble model in predicting the distress status of firms. Among the high-frequency firm-level indicators, CPF Late Payments and Retrenchment Notifications have higher variable importance scores than FWL Defaults, JTC Rental Arrears and Electricity Arrears [Exhibit 4]. This suggests that labour costs are a more significant component of business costs for firms in most sectors of the economy, relative to other costs such as utilities and rent.[8] As such, firms that are unable to meet their obligations to their workers are more likely to be in distress.

Exhibit 4: Variable Importance of Selected Variables for the Ensemble Model



7 *Variable importance scores measure the salience of each input variable to the model when predicting the target variable for each observation. For technical details on the calculation of variable importance, refer to Zhu et al. (2015).*
8 *For more information on business cost conditions in Singapore, see Tan (2022).*

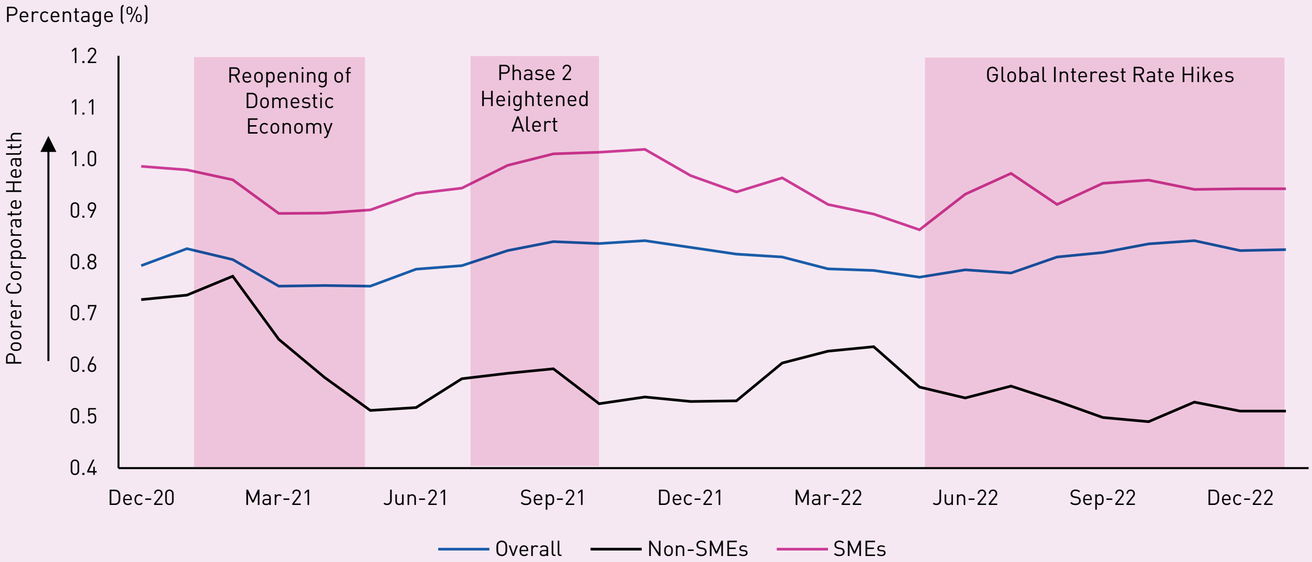
TRENDS IN CORPORATE HEALTH AND CHI

Based on ACRA's data, we observe that corporate health in Singapore, as measured by the proportion of firms in distress within a three-month forward-rolling window, is affected by macroeconomic developments [Exhibit 5].

For example, the proportion of firms in distress fell from 0.83% at the start of 2021 to 0.75% in March 2021, in line with the lifting of social restrictions and reopening of the domestic economy in Phase 3 of the COVID pandemic era. However, the institution of Phase 2 Heightened Alert measures in July 2021 led to a steady increase in the proportion of firms in distress, eventually reaching a peak of 0.84% in November 2021, one month after the transition to Stabilisation Phase. As the economy subsequently transitioned to a more relaxed social restriction profile, the proportion of firms in distress fell to 0.77% by May 2022. Thereafter, global events such as the hike in global interest rates led by the US Federal Reserve resulted in a reversal of trends, with the proportion of firms in distress rising to reach 0.82% in January 2023.

Breaking down by firm types, we observe that the proportion of firms in distress among SMEs is systematically higher than that among non-SMEs. The hike in global interest rates in the second half of 2022 also had a greater impact on SMEs than on non-SMEs, although its impact appeared to have tapered off in early 2023.

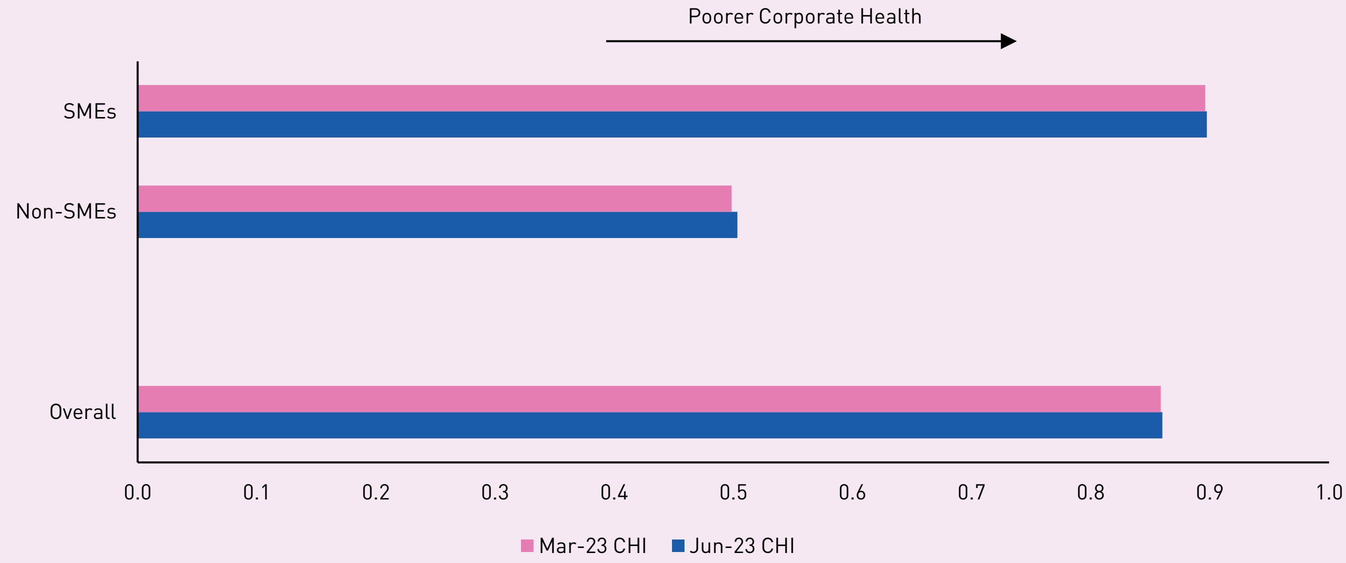
Exhibit 5: Percentage of Firms in Distress in the Overall Economy and by Firm Size, Actualised Data



Note: The actual percentage of firms in distress (based on entity status information from ACRA) is computed within a three-month forward-rolling window. For example, the data point for January 2023 shows the actual percentage of firms in distress in the months of January to March 2023.

Leveraging on the CHI model to obtain a forward-looking assessment of corporate health in Singapore, we observe that the latest reading of the CHI for the overall economy in June 2023 (i.e., for the three-month period of June to August 2023) remains stable, suggesting that overall corporate health is likely to be stable in the near term [Exhibit 6]. Similar to the overall economy, the CHI readings for SMEs and non-SMEs in June 2023 also suggest that their corporate health is likely to remain resilient in the near term.

Exhibit 6: Recent Trends in the CHI by Firm Size

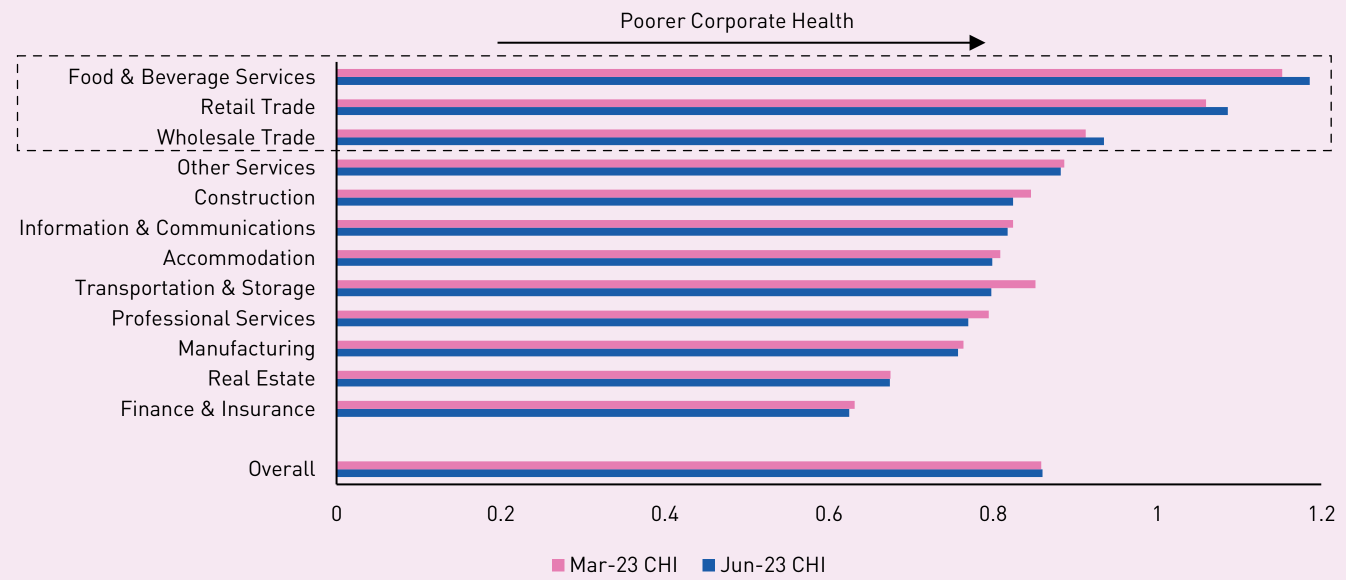


Note: TThe CHI shown for each month represents the predicted probability of firms being in distress within a three-month period. For example, the CHI for June 2023 represents the predicted probability of firms being in distress in the months of June, July and August 2023.

Breaking down by sectors, we find that the latest June 2023 CHI readings for most sectors have either remained similar or are lower than the readings in March 2023 [Exhibit 7]. Among the domestically-oriented sectors, the only exceptions are the food & beverage services and retail trade sectors. Notably, their respective June 2023 CHI readings ticked up to 1.2% and 1.1% from their March 2023 CHI readings, suggesting that corporate health in these sectors could weaken slightly in the near term. This in turn may reflect the effect of higher business costs on SMEs in the sectors. Nonetheless, the corporate health situation in these sectors is likely to be better as compared to the years before the COVID-19 pandemic, when the actual proportion of firms in distress in the food & beverage services and retail trade sectors averaged 1.6% and 1.5% respectively in 2015-2019.[9]

Among the outward-oriented sectors, there was a slight increase in the June 2023 CHI reading for the wholesale trade sector to 0.9% compared to its reading in March 2023. While this likely reflects the impact of external headwinds faced by firms in this sector arising from weak global demand, the June 2023 CHI reading remained lower than the actual proportion of firms in distress in the wholesale trade sector in 2015-2019, which averaged 1.3%.[9]

Exhibit 7: Recent Trends in the CHI by Sectors



Note: The CHI shown for each month represents the predicted probability of firms being in distress within a three-month period. For example, the CHI for June 2023 represents the predicted probability of firms being in distress in the months of June, July and August 2023.

9 *The actual proportion of firms in distress for each sector over the period of 2015-2019 is computed as follows. First, for each month across the years of 2015 to 2019, compute the actual proportion of firms in distress in the sector over a three-month forward-rolling window. Second, average the computed proportions over all the months in the five-year period of 2015-2019.*

CONCLUSION

In this study, we have constructed a model to measure corporate health using predictive machine learning techniques (i.e., the CHI model), to facilitate the timely surveillance of corporate health at the overall economy level, as well as by sectors and firm types. Indicators that provide high informational value include ACRA financial variables such as total non-current assets, net profit after tax and total liabilities, as well as high-frequency firm-level indicators such as CPF Late Payments and Retrenchment Notifications.

Looking ahead, our CHI model suggests that overall corporate health in Singapore is likely to remain stable in the near term. While the risk of distress has edged up for firms in selected sectors such as food & beverage services and wholesale trade, possibly due to higher business costs and weak global demand respectively, their latest CHI readings remain low compared to the proportion of firms in distress in the years before the COVID-19 pandemic COVID period. We will continue to monitor the trends in CHI closely in the coming months given the prevailing economic headwinds.

Contributed by:

Mr Ong Chong An
Senior Economist
Economics Division
Ministry of Trade and Industry

Mr Rafi Kamsani
Economist
Economics Division
Ministry of Trade and Industry

REFERENCES

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance 23 (4)*, 589 - 609.

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction.*Expert Systems with Applications 83* , 405 - 417.

Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.

Graczyk, M., Lasota , T., Trawiński , B., & Trawiński , K. (2010). Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal. *In Intelligent Information and Database Systems: Second International Conference, ACIIDS*, Hue City, Vietnam, March 24 26, 2010. Proceedings, Part II 2 (pp. 340 - 350). Springer Berlin Heidelberg.

Liang, D., Tsai, C. F., Dai, A. J., & Eberle, W. (2018). A novel classifier ensemble approach for financial distress prediction. *Knowledge and Information Systems 54* , 437 - 462.

Liang, D., Tsai, C. F., Lu, H. Y. R., & Chang, L. S. (2020). Combining corporate governance indicators with stacking ensembles for financial distress prediction. *Journal of Business Research 120* , 137 - 146.

Olson, D. L., Delen , D., & Meng, Y. (2012). Comparative analysis of data mining methods for bankruptcy *Decision Support Systems 52 (2)*, 464 - 473.

Sun, J., Fujita, H., Chen, P., & Li, H. (2017). Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble. *Knowledge Based Systems 120* , 4 - 14.

Tan, Y. L. (2022). Business Cost Conditions in Singapore's Manufacturing and Services Sectors. *Economic Survey of Singapore 2022*, 27 - 34.

Tsai, C. F., Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing 24* , 977 - 984.

Tsai, C. F. (2014). Combining cluster analysis with classifier ensembles to predict financial distress. *Information Fusion 16* , 46 - 58.

Zhu, R., Zeng, D., Kosorok, M. R. (2015). Reinforcement Learning Trees, *Journal of the American Statistical Association*. 110 (512), 1770-1784.