The impact of Income Shocks to Probability of Default Estimates

Credit Scoring and Credit Control XVII, Aug 2021

Vassilis Ioannou

Head of Data Science,
Equifax UK

Raffaella Calabrese

Senior Lecturer in Data Science,
Business School University of Edinburgh

Introduction : Probability of Default & Income Shocks

> **Credit scores** have been long used for **ranking the risk of loans** to support credit strategies and as an **input to probability of default estimates** for loss provisioning purposes.

> Credit scores traditionally reflect information available for wide segments of the scored population such as **repayment** and **delinquency variables**, **credit card utilization levels**, **hard-searches** and a long list of related characteristics historically available either internally or via a credit bureau.

> Despite it's straightforward linkage to one's ability to meet financial obligations and repay loans on time, **Income information** had only had marginal role in automated credit risk prediction and usually in the form of **declared Income** (confirmed or not).

Introduction : Probability of Default & Income Shocks

> Banks with a substantial current account customer base also had the possibility to derive Income, or correlated variables, based on turnover information to improve their predictions and better support their clients' needs.

> Open Banking and the **Current Account Turnover Data** (**CATO**) in the UK provide additional information access opportunities, but not without challenges compared to traditional Bureau data:

> Limitations on the usage/handling of such data (Open Banking, CATO)
>
> One shot or 90 days recurring availability (Open Banking)
>
> Excluding the un-consented part of the population (CATO)
>
> In many cases only proxies of Income are realistically available

Introduction : Probability of Default & Income Shocks

> Despite the difficulties, the COVID crisis emphasized anew the importance of Income information and especially the impact of **income streams' shocks** to credit risk.

> During the early stages of the crisis **Governments** intervened with programs that timely and effectively **supported households facing abrupt income disruption**, successfully (so far) averting a bad debt credit crisis as well.

> Better identification of income shocks on an individual basis would allow banks to better support their clients at the time they need it most.

> Improved quantification would also lead to more accurate credit loss estimates, which is a constant goal for financial institutions.

Methodological approach

> The current presentation discusses an approach that can be used to estimate **updated probabilities of default**, conditional to traditional credit scores/variables and loan lifecycle dynamics, when a **probable income-shock event is identified.**

> To identify such probable income shock events, Equifax characteristics (features) available to member lenders, have been used to develop an **Income-shock proxy.**

> For the reasons already mentioned, this study makes use of **simulated data** - not representative of any specific lender or the UK average population, but **developed in a way that makes them relevant to UK portfolios that would have a similar composition to the simulated data sample**.

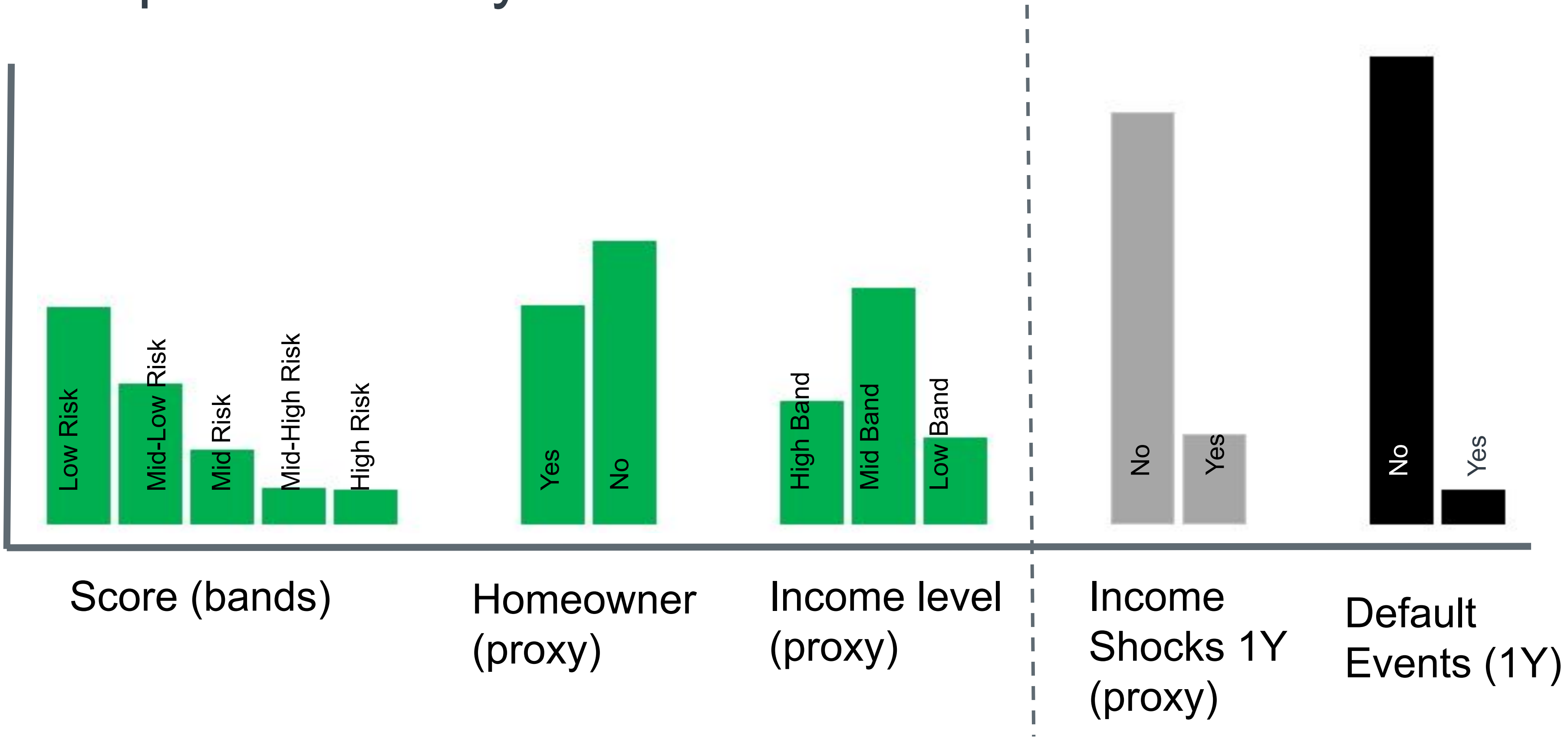
Methodological approach

> The data used for this study represent personal loans originated during the period of one year, followed by 2 additional years of performance observation

> They reflect a period of normal business conditions and they do not include post-covid observations or performance

> This sample is used to evaluate the significance of the income-shock proxy and develop a survival analysis model that can be used to simulate the impact of such events after loan origination

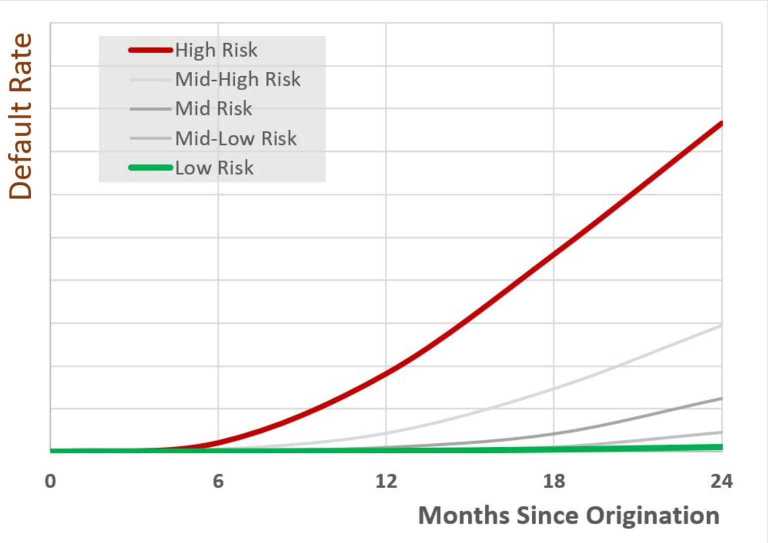
Composition % by Risk Driver



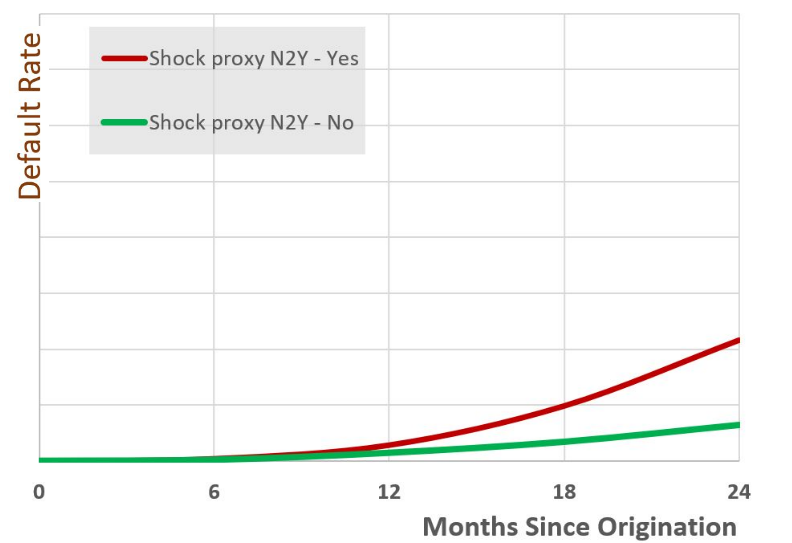
Cumulative Incidence Functions (default event)

Bureau Scores are essentially behavioral models, capturing a wide range of the consumer behavior, and are strongly correlated to default rates.

Bureau Score (banded)



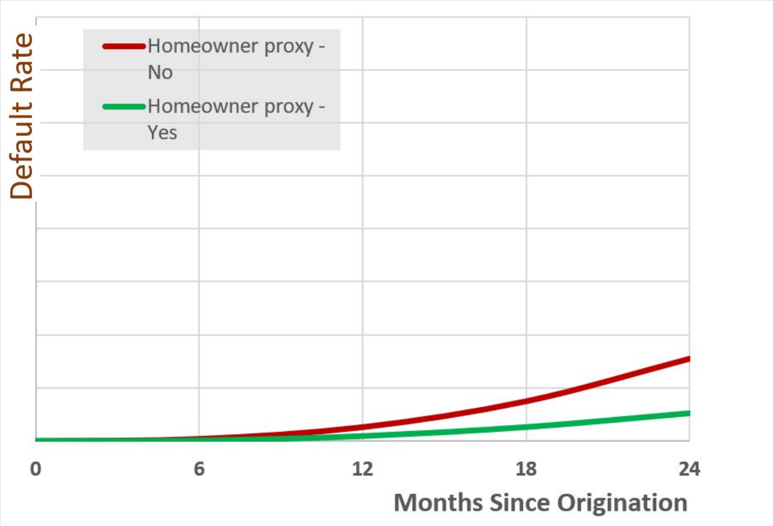
Income Shock proxy N2Y



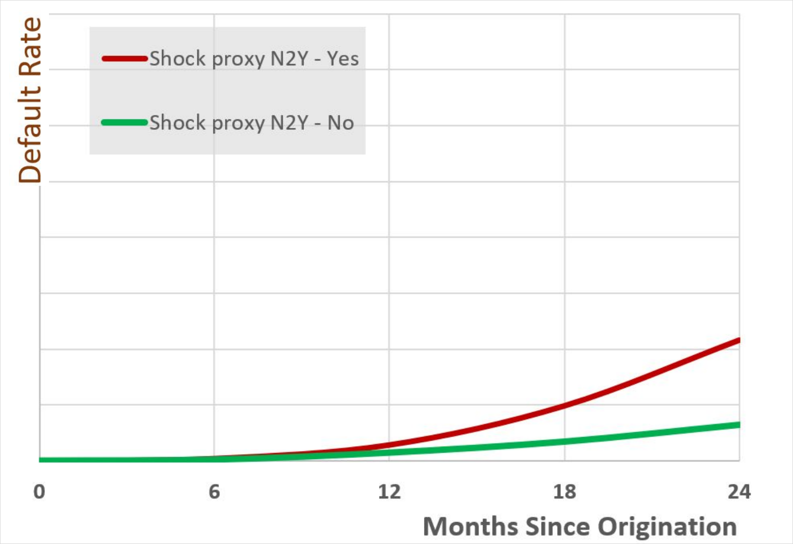
Cumulative Incidence Functions (default event)

Loans for which the Income Shock flag is triggered during the next 2 years have higher default rates; the significance was also tested within a multivariate model

Homeowner flag (proxy)



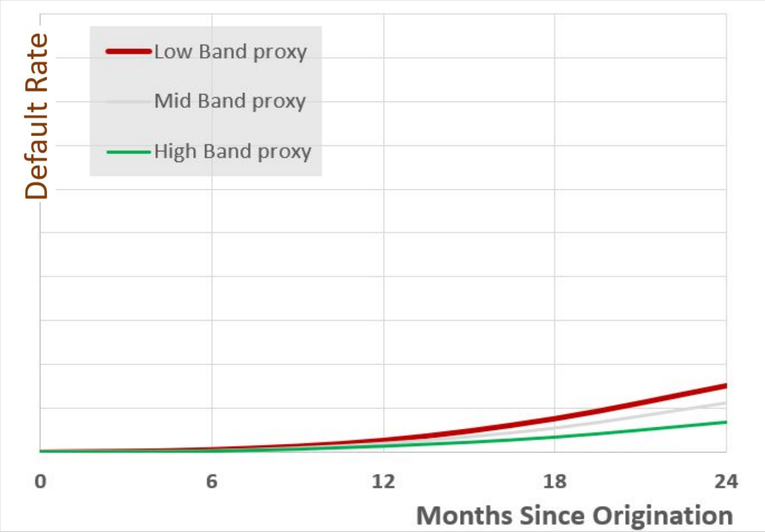
Income Shock proxy N2Y



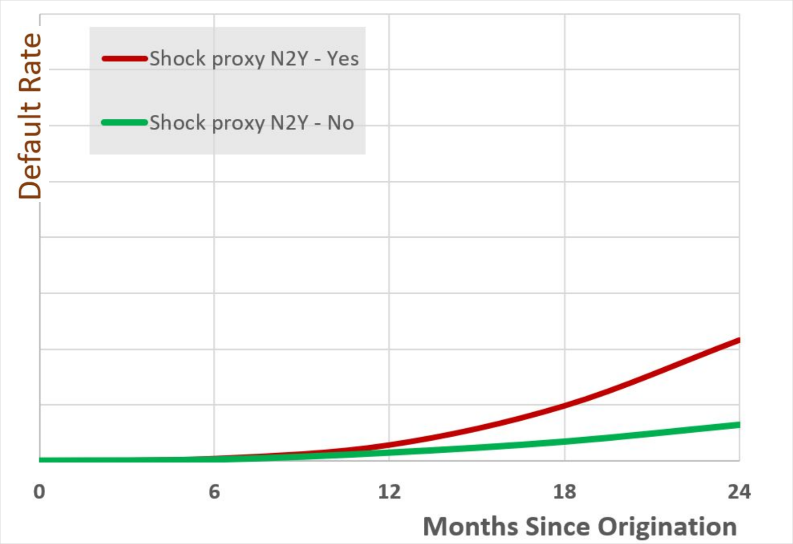
Cumulative Incidence Functions (default event)

The income shock flag has a much stronger association with default compared to the income level flag at origination.

Income level flag (proxy)



Income Shock proxy N2Y



Survival Analysis Model

Discrete-time survival analysis, fitting a multinomial logit model to predict default in the presence of the competing risk of prepayment.

$$\ln\left(\frac{P_{i1t}}{P_{i0t}}\right) = \beta_1{}'X_{it}$$

$$\ln\left(\frac{P_{i2t}}{P_{i0t}}\right) = \beta_2{}'X_{it}$$

P_{i0t} : Probability of loan i surviving time t

P_{i1t} : Probability of loan i defaulting during time t

P_{i2t} : Probability of loan i being repaid during time t

Semiannual time intervals, $t = 1,2,3,4$

The cumulative incidence function used in later stages of the analysis, where cumulative PD rates are required, is calculated considering predicted probabilities for both events, default and prepayment.

Covariates and Odds Ratio Estimates (default event)

Covariates	Time-varying
Score at origination (bands)	Coefficients only
Homeowner proxy flag (at origination)	No
Loan Age/Loan Term (transformed)	Covariate only
Income Shock proxy flag (current period)	Covariate only
Income Shock proxy flag (previous period)	Covariate only

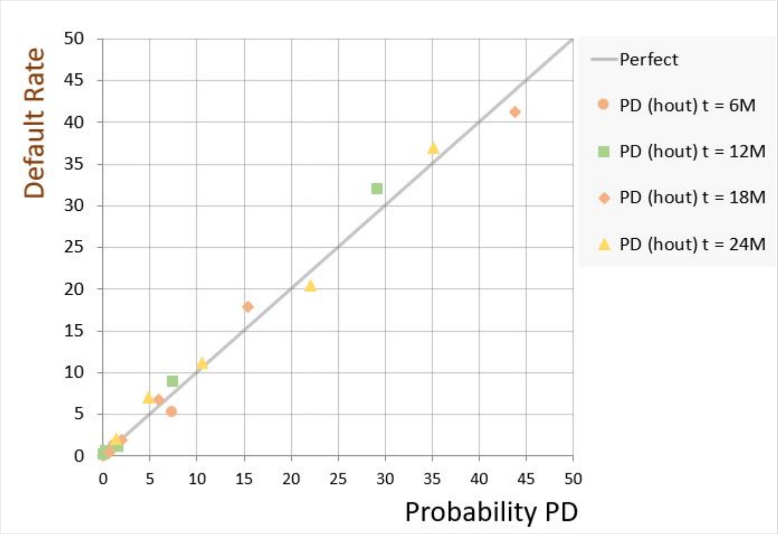
Covariates	Odds Ratio Estimates 95% Confidence Intervals	
Income Shock proxy flag (current period)	2.7	3.3
Income Shock proxy flag (previous period)	1.8	2.3

Income Shock proxy is found to be statistically significant and important within the multivariate model.

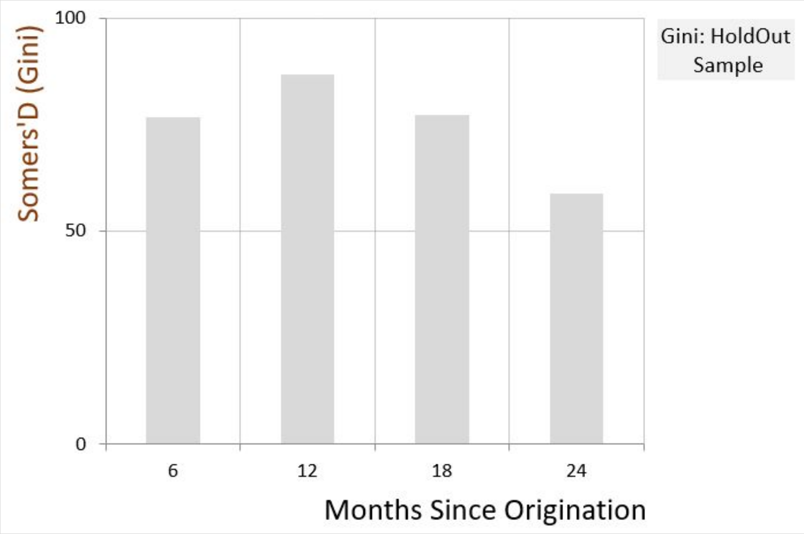
Model fit to the data (default event)

The model has a good fit to the data (actual vs predicted), and ranking performance when evaluated using the hold out sample.

Model fit to the data



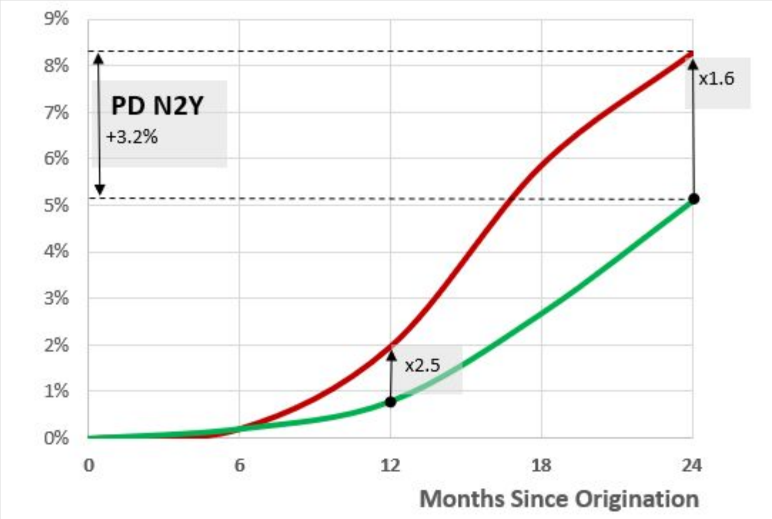
Ranking Performance



Simulation at the loan level (default rates)

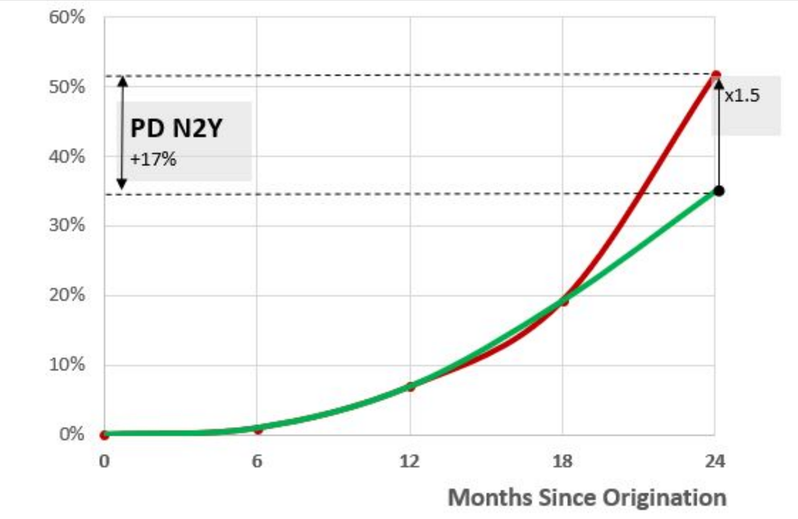
This type of model can be used to simulate the effect of the income-shock flag raised at different points during the life of loans with different risk profile.

Shock flag raised 6-12M, mid-level risk loan



Homeowner proxy: yes, Loan Term: 24M

Shock flag raised 18-24M, high risk loan



Homeowner proxy: no, Loan Term:60M

Conclusions

- New sources of data such as CATO and Open Banking allow to capture income-related information, and can be used to support the identification of probable shock events

- Synthetic/Simulated Data can be used to promote credit research tackling barriers in the use of actual sensitive data

- The proposed approach allows for the quantification of income shocks to default risk, at an individual loan level and at a portfolio level following a bottom-up approach

- Encouraging results support further research using additional methodologies and testing on specific portfolios/real data (always compliant to applicable regulations)