BANK OF CANADA
BANQUE DU CANADA

Technical Report/Rapport technique—122

Last updated: October 3, 2022

Forecasting Banks' Corporate Loan Losses Under Stress: A New Corporate Default Model

by Gabriel Bruneau, Thibaut Duprey and Ruben Hipp

Financial Stability Department
Bank of Canada
gbruneau@bankofcanada.ca
tduprey@bankofcanada.ca
rhipp@bankofcanada.ca

DOI: https://doi.org/10.34989/tr-122 | **ISSN** 1919-689X

Acknowledgements

The authors would like to thank Grzegorz Halaj, Guillaume Ouellet Leblanc, Xiangjin Shen and an anonymous referee for their valuable comments. This work has benefited greatly from outstanding research assistance by Vatya Kishore. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada.

Abstract

We develop a corporate default model to forecast corporate loan losses of the Canadian banking sector under stress. First, we tackle a data gap by reconstructing historical default probabilities for banks' loan portfolios. Second, we estimate tail elasticities to capture non-linear relationships between macrofinancial conditions and default probabilities. By explicitly modelling default probabilities associated with macroeconomic tail events, this model significantly improves the Bank of Canada's stress-testing infrastructure.

Topics: Economic models; Financial institutions; Financial stability; Financial system regulation and policies

JEL codes: C22, C52, C53, G17, G21, G28

1 Introduction

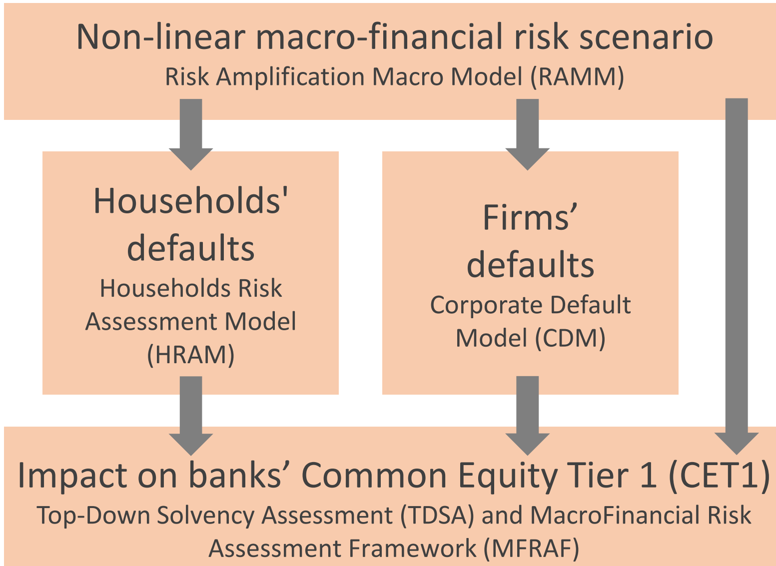
Regulators routinely perform stress tests of financial institutions to assess the resilience of banks' balance sheets under extreme but plausible shocks, typically a severe recession. One of the main sources of losses for banks comes from credit risk—the risk that borrowers fail to meet their contractual obligations. Unexpected credit losses reduce banks' capital directly, while expected credit losses generate higher loss provisions, thereby limiting banks' ability to deploy their capital to the real economy. In both cases, large credit losses could impair banks lending ability and, in turn, the functioning of the broader financial system.

In this report, we present a new credit risk model for the corporate loan portfolio of the Canadian banking sector. The new Corporate Default Model (CDM) maps relevant macroeconomic stress factors into aggregate and industry-specific default probabilities. We use the model to forecast possible default probability paths under extreme but plausible macrofinancial risk scenarios. The new model improves on the previous corporate credit risk model described in the Framework for Risk Identification and Assessment (FRIDA) in section 3.2 of MacDonald and Traclet (2018). FRIDA is a suite of models used to conduct top-down macrofinancial stress tests at the Bank of Canada (see Figure 1). The Risk Amplification Macro Model (RAMM) is the FRIDA module that generates the macrofinancial risk scenarios that are fed into the CDM module. We then use the implied corporate default probabilities in conjunction with the household loan defaults obtained through the Household Risk Assessment Model (HRAM) to compute scenario-consistent credit losses and assess the capital shortfall on banks' balance sheets.[1]

The new CDM improves corporate default probability modelling in four key dimensions. First, we implement a novel filter for extracting default probabilities—the variable of interest—from the stocks of total loans and impaired loans, which are the variables available with sufficient coverage and industrial disaggregation. We can

[1]More details on FRIDA are in MacDonald and Traclet (2018); on RAMM, in Tuzcuoglu (forthcoming); on HRAM, in Peterson and Roberts (2016); and on MFRAF, in Fique (2017a).

Figure 1: Framework for risk identification and assessment—an overview



then directly estimate the historical relationship between default probability and key macrofinancial stress factors.

Second, we begin by modelling the aggregate default probability before providing an industry decomposition anchored around that aggregate probability. This approach is more robust to outliers and ensures consistency between simulations of aggregate and industry-specific probabilities of default.

Third, we broaden the model's set of potential macrofinancial stress factors by including, for example, risk factors associated with the foreign exposures of Canadian banks. To avoid overfitting, we use variable selection and cross-validation to maximize the model's out-of-sample performance, resulting in a parsimonious specification that can replicate historical peaks in default probabilities.

Fourth, through a horse race between modelling techniques, we determine that quantile regressions are the best at capturing the non-linearities that might exist in adverse macrofinancial risk scenarios. The previous generation of the model, in contrast, relied on linear regressions.

Our new corporate default model shows strong in-sample and out-of-sample performance.[2] In simulation exercises, the linear version of the model can already reproduce historical levels of probabilities of default, but with relatively narrow confidence bands. In contrast, our non-linear model can generate larger probabilities of default because we can rely on a stronger response in the tail of the distribution and a longer duration of the tail event. This increased flexibility in crisis simulation allows us to generate counterfactual probabilities of defaults that are more severe than the peaks observed in our dataset. This is particularly important given that Canada has not recently observed a severe banking crisis, so we would otherwise underestimate risks when conducting stress tests.

The remainder of the paper is organized as follows. Section 2 discusses the implementation of our new methodology to filter probabilities of defaults. Sections 3 and 4 present our modelling strategy and report our estimation and simulation results both for the aggregate corporate sector and by industry. An application is shown in section 5. Finally, section 6 concludes.

2 A new method to filter probabilities of default

Default probabilities are a key component of banks' credit risk. Specifically, the expected loss EL_t in period t can be expressed as:

$$EL_t = PD_t \cdot EAD_t \cdot LGD_t,$$

where PD_t, EAD_t and LGD_t represent the probability of default, exposure at default and loss given default, respectively. In this report, we focus on modelling and predicting probabilities of defaults (PDs). We extract PDs that are useful for our purposes from partial information because we do not observe them directly (except in aggregate portfolios or in short time series from public reports).

In the past, several approaches tried to fill these gaps. For example, Misina, Tessier,

[2] Note that we estimate the model until 2019 to avoid the COVID-19 period, which has been characterized by loan deferrals and extraordinary government support that distorted the statistical relationship between default probabilities and economic conditions.

and Dey (2006) and Djoudad and Bordeleau (2013) developed measures of default probabilities by industry, based on bankruptcy rates. Unfortunately, because bankruptcy is a narrow measure of default, these approaches do not explain expected losses. Moreover, because bankruptcies are the final stage in a default, they are reported with a lag. Covas, Rump, and Zakrajšek (2014) used charge-offs, which are not observed over a long enough time horizon for our purposes.

To quantify PDs, we use data on impaired loans because impaired loans data is available over a long sample and available with an industry disaggregation. These features enable us to estimate the relationship of PDs to macroeconomic conditions. For example, the previous version of the CDM used the gross impaired loan ratio as a proxy for the dynamics of PDs. An ad hoc formula converted the impaired loan ratio to probabilities by accounting for a lag between the impairment and the loss of a loan. However, we want to improve on the interpretation of this proxy to allow for more rigorous non-linear models.

Thus, in this section, we present a novel filter for extracting probabilities of defaults —the variable of interest—from the stock of total loans and impaired loans—the variables with sufficient coverage and industry disaggregation.

2.1 Method

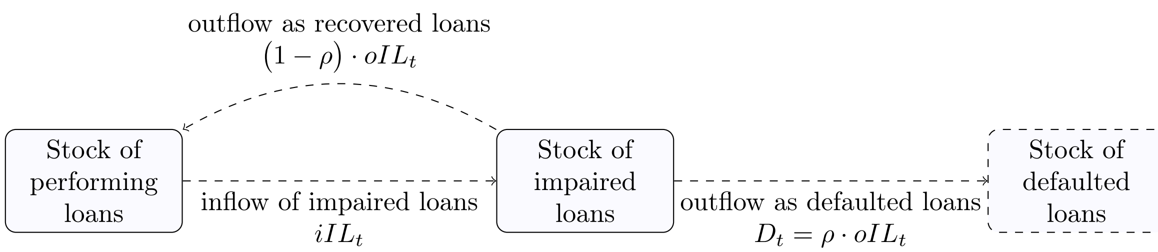
We define PD_t as the (expected) share of the total loans that generate losses in period t. That is:

$$PD_t = \frac{D_t}{TL_t}, \tag{1}$$

where D_t denotes the loans that defaulted in period t, and TL_t denotes the stock of total loans in the same period. This definition differs slightly from credit-level calculations in that it considers the entire loan portfolio of a given bank and sector. This assumption is critical because there is insufficient data to estimate the number of defaulted loans, and the probability of default cannot easily be derived from the total number of defaulted loans.

In line with the definition from the Bank for International Settlements (Basel Committee on Banking Supervision 2006), we understand an impaired loan as a loan that is predicted to have *worse* future cash flows than initially expected. Given that the worsening of expectations comes first, we assume that each loan that eventually defaults goes through the stage of being impaired first. Figure 2 illustrates the concepts of flow and stock and introduces inflow iIL_t and outflow oIL_t of the impaired loan stock. The figure depicts stock of loans by boxes, and flows of loans per period by arrows. Dashes indicate that the number is not observed. This figure illustrates two problems in deriving PDs from IL_t. First, we define the unknown share of loans that go from being impaired to defaulting as ρ, and the respective cure rate as $1-\rho$. Second, iIL_t and oIL_t are unobserved flow variables, while only the quarterly stock of impaired loans IL_t is available. In what follows, we show how we derive the inflow and outflow from the observed stock.

Figure 2: Illustration of flow and stock concepts



Determining the PD path. We derive inflow and outflow from the differences in the stock of impaired loans:

$$\Delta IL_t = iIL_t - oIL_t = IL_t - IL_{t-1}. \tag{2}$$

We assume that each loan in this stock faces an outflow probability that depends solely on the timing of entering the stock. For example, assume a loan gets impaired in period t; then it faces a positive outflow probability $f_\lambda(k)$ in period $t+k$. In this report, we use the Poisson distribution with mean λ.[3] Formally, $f_\lambda(k)$ is the expected

[3]More generally, this method is applicable to any probability distribution with positive support.

share of outflowing loans k periods after being impaired. For the set of loans, the outflows can be written as:

$$oIL_t = \sum_{k=0}^{\infty} f_\lambda(k) \cdot iIL_{t-k}.$$

By rearranging equation (2), we can calculate the inflows and thus also outflows recursively:

$$\begin{aligned} iIL_t &= \Delta IL_t + oIL_t = \Delta IL_t + \sum_{k=0}^{\infty} f_\lambda(k) \cdot iIL_{t-k}, \\ &= \frac{1}{1-f_\lambda(0)}\Big(\Delta IL_t + \sum_{k=1}^{\infty} f_\lambda(k) \cdot iIL_{t-k}\Big). \end{aligned}$$

Since in practice we have to provide an initial value of inflow, we impose some boundaries on these numbers. This step makes the results more tractable and robust. By definition, the outflow cannot be larger than the observed negative difference plus the current inflow, and the inflows cannot be negative. That is:

$$oIL_t = \min\{\sum_{k=0}^{\infty} f_\lambda(k) \cdot iIL_{t-k}, \quad iIL_t - \Delta IL_t\}, \tag{3}$$

$$iIL_t = \frac{1}{1-f_\lambda(0)}\max\{\Delta IL_t + \sum_{k=1}^{\infty} f_\lambda(k) \cdot iIL_{t-k}, \quad 0\}. \tag{4}$$

Equations (3) and (4) now recursively define the outflow and inflow of impaired loans. From equation (1) and Figure 2, we know that $PD_t = \frac{\rho \cdot oIL_t}{TL_t}$.[4] The percentage change of PDs therefore reads:

$$pPD_t = \frac{PD_t}{PD_{t-1}} = \frac{\rho \cdot oIL_t \cdot TL_{t-1}}{\rho \cdot oIL_{t-1} \cdot TL_t} = \frac{oIL_t \cdot TL_{t-1}}{oIL_{t-1} \cdot TL_t}. \tag{5}$$

We tried alternative specifications, but we do not observe data to formally assess one distributional family against another. Still, the shape of the distribution lines up with expert judgement and aggregate moments.

[4]We assume ρ to be an accounting-specific parameter that does not change over time. That is, banks maintain the same accounting standards during times of stress, and changes in PDs are fully attributable to changes in impaired loans.

As a result, equations (3) and (5) back out pPD_t with inputs ΔIL_t and TL_t and a certain level of outflow in period $t = 0$. The level of the PD, however, is still undetermined and needs to be anchored.

Anchoring the PD path. To determine the level of PD, we use a shorter time series of observed PDs—PD_t^{anchor} from $t = t^*, ..., T^*$—as an anchor for the level. This anchor is used as the dependent variable in the regression

$$\underbrace{PD_t^{anchor}}_{y_t} = \beta \cdot \underbrace{\prod_{\tau=t^*}^{t} pPD_\tau}_{x_t} + \varepsilon_t, \quad \forall t = t^*, ..., T^*$$

estimated by the OLS. Given the new starting point $\hat{\beta} = \widetilde{PD}_{t^*-1}$, we can develop the PD path forward and backward, respectively:

$$\widetilde{PD}_t = \hat{\beta} \cdot \prod_{\tau=t^*}^{t} pPD_\tau, \quad for \quad t = t^*, ..., T,$$
$$\widetilde{PD}_t = \hat{\beta} \cdot \prod_{\tau=t}^{t^*-2} pPD_\tau^{-1}, \quad for \quad t = 1, ..., t^* - 2.$$

In a nutshell, the series $\widetilde{PD}_t$ is the full sample PD with dynamics determined by pPD_t and levels determined by PD_t^{anchor}.

Initializing the filter. The previously described filter works recursively and requires calibration of the initial inflow and outflow, which determine the size of the average flow per period. While the size of the outflow gets cancelled out in equation (5), it is important in the calculation of equations (3) and (4) because it determines the sensitivity of the lower bound of inflows (i.e., all inflows must be positive). For example, if the first inflow is set too large, a high inflow of impaired loans is possible even with a strongly negative ΔIL_t. As a result, we start the filter at the first tranquil period and set the initial inflow and outflow equal to their sample averages.

To do so, we implement the following algorithm:

1. Find $\mathcal{O} = \text{argmin}_{t>3} t \cdot \mathbb{1}\left(\frac{1}{3}\frac{IL_t}{TL_t} > \frac{1}{3}\frac{IL_{t-3}}{TL_{t-3}}\right)$ to determine the first three-quarter moving average minimum in the sample.

2. Set the initial relative outflow α to $\alpha = 0.5$.

3. Set the inflow and outflow of period $\mathcal{O}$: $iIL_{\mathcal{O}} = oIL_{\mathcal{O}} = \frac{\alpha}{1-\alpha} \cdot IL_{\mathcal{O}}$.

4. Calculate the subsequent inflow and outflow with the recursive formula in equation (3).

5. Calculate the resulting relative outflow with $\alpha^* = T^{-1} \sum_t \frac{oIL_t}{(IL_t + iIL_t)}$.

6. If $|\alpha - \alpha^*| > 1e - 8$, set $\alpha = \alpha^*$ and repeat steps 3 to 6. Otherwise, end the algorithm.

We backfill all values before the initialization, $t < \mathcal{O}$, with the percentage growth rate $pPD_t = \frac{IL_t \cdot TL_{t-1}}{IL_{t-1} TL_t}$, which works reasonably well because we set $\mathcal{O}$ in a way that PDs decrease before it.

2.2 Data

We use two different regulatory returns of domestic systemically important banks (D-SIBs) for the calculation of PDs:[5] (1) A2 for dynamics[6] and (2) RAPID2 for levels (also known as BF regulatory return).[7] First, beginning in the second quarter of 1994, A2 reports impaired loans and total loans per quarter by industry for Canadian D-SIBs.[8] This return allows us to determine the dynamics of PDs for each bank and industry separately.

Second, RAPID2 is a granular loan-level dataset that collects business loans worth more than Can$10 million.[9] It is available only from the first quarter of 2014 onward,

[5]The share of D-SIB loans to total bank loans is over 90%.

[6]See Non Mortgage Loans (A2) of the Office of the Superintendent of Financial Institutions.

[7]See Technical Specification (BF) of the Office of the Superintendent of Financial Institutions.

[8]We use the following industries: 1) natural resources, 2) mining, 3) manufacturing, 4) construction and real estate, 5) transportation, communications and utilities, 6) wholesale trade, 7) retail trade, 8) services, 9) financial and 10) others. More details are provided in Appendix A. Appendix B outlines extensions of our framework to the non-corporate sector, namely consumer loans, residential mortgages and non-residential mortgages.

[9]RAPID2 provides time-series and cross-sectional data but does not cover loan-level data for

so this shorter time series is used only as a level anchor when converting the impaired-loans data from A2 into probabilities of defaults. The RAPID2 dataset contains the borrower's one-year probability of default as reported by the financial institution. We convert it to quarterly levels using the formula:

$$PD^q_{i,t} = 1 - (1 - PD^y_{i,t})^{\frac{1}{4}},$$

where the superscripts q and y describe quarterly and yearly default probabilities, respectively. Then we aggregate these probabilities to industry PDs by calculating the weighted average within an industry. The PD for industry ind, which is then used as the anchor for this industry-specific PDs, is:

$$PD^{anchor,ind}_t = \sum_{i \in ind} w_{i,t} PD^q_{i,t}$$
$$\text{with} \qquad w_{i,t} = \frac{L_{i,t}}{\sum_{j \in ind} L_{j,t}},$$

where $L_{i,t}$ refers to the outstanding Canadian-dollar amount of loan i.

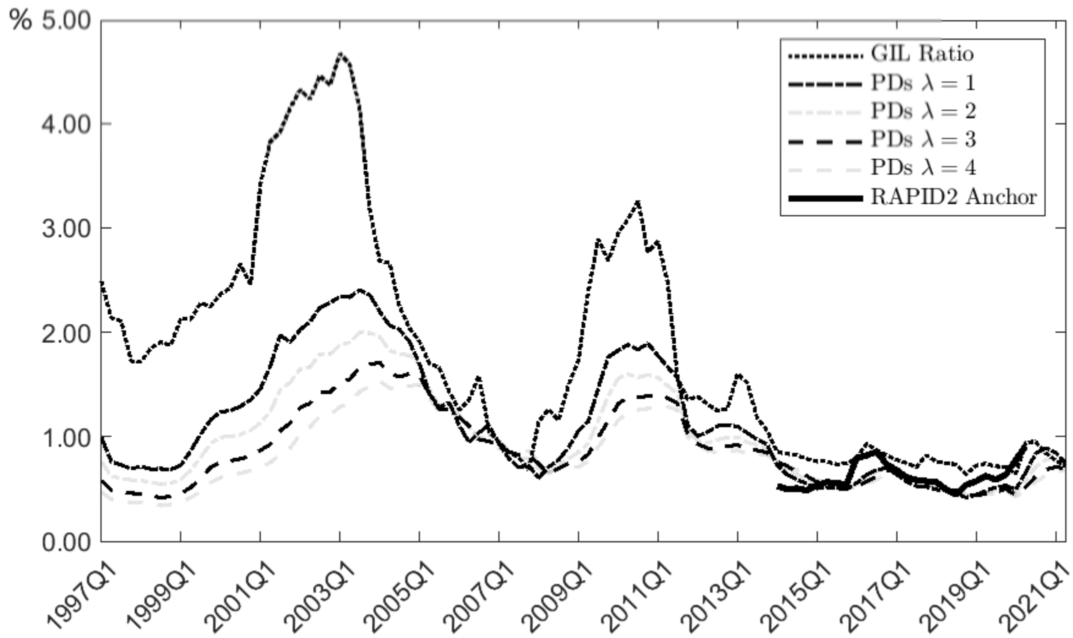
The last free parameter in the filter is λ, the average stay of a loan in the impaired-loan bucket. This parameter is mainly responsible for the smoothness of the filtered probability of default and does not change its dynamics drastically. We set λ to a value between one- and three-quarters based on Monte Carlo simulations of the impaired-loan dynamics and on expert opinions, respectively. In the following section, we present empirical results of the filter and investigate its robustness to the choice of λ.

2.3 Results

Figure 3 shows the aggregated PD series for $\lambda = \{1, 2, 3, 4\}$, anchored to the shorter PD series of RAPID2 over the first quarter of 2014 to the fourth quarter of 2019. The dotted black line is the ratio of gross impaired loan (GIL) from the A2 dataset, and

smaller business loans; our results, therefore, might be biased downward if larger loans to larger companies have systematically lower PDs.

Figure 3: Historical default probabilities—aggregated corporate sector, all D-SIBS



the other lines are converted PDs with different λ. Finally, the thick bold line is the anchor PD series as reported in the RAPID2 return. When this ratio is compared with the new calculations of PDs, two things become clear. First, the level of the converted PDs (for $\lambda = 1$) is about half of the GIL ratio. The anchoring of the RAPID2 dataset, which reports PDs lower than the GIL ratio, is largely responsible for this result. Second, the longer the loan is assumed to stay in the impaired bucket, as parameterized by λ, the longer it takes before an impaired loan possibly defaults. Thus the peak of the probability of default is delayed relative to the peak of the GIL ratio, and increasingly so as λ increases.[10]

3 Modelling aggregate probability of default

We now turn to the modelling of aggregate PD (i.e., corporate sector–wide PD) under various macrofinancial conditions. In this section, we present the macrofinancial

[10] In turn, the shock that triggers loan impairment usually precedes the peak in the GIL ratio. Therefore, it is to be expected that a macrofinancial shock would predict relatively well subsequent peaks in the probability of default. Given that losses take time to materialize on banks' balance sheets, a delay from the impaired-loans ratio to the PDs is desired.

factors used to explain PDs and discuss how they were chosen for inclusion in the final model. We take an agnostic approach, examining a wide range of potential explanatory variables and keeping only those that contain the most relevant information. We also present estimation and simulation results for the two chosen estimation methodologies, an autoregressive model and a quantile regression model, which were chosen after a horse race of linear and non-linear modelling techniques.[11] Finally, we present backtesting of historical PDs.

3.1 Which macrofinancial factors predict PDs?

Table 1 contains a complete list of the 25 macrofinancial factors considered as potential explanatory variables of PDs, as well as their conventional signs.[12] This selection includes a wide range of domestic and foreign macroeconomic and financial explanatory variables. We first collected the main drivers of credit risk identified in the literature[13] and then selected those variables that are part of our macroeconomic models (like RAMM) used to generate macrofinancial risk scenarios.

The 25 potential explanatory variables are divided into several categories. First, the economic and financial cycle has an impact on loan quality. Deteriorating economic and financial conditions, such as slow economic growth, rising unemployment, falling asset prices or a sudden tightening of financial conditions, are associated with debt

[11] We present only the most relevant linear and non-linear models, but we also tested alternative specifications: the vector autoregression model by modelling all industries jointly, but it did not beat univariate models out of sample; the vector error correction model, but probabilities of default do not exhibit unit roots; univariate and multivariate autoregressive Markov-switching models, but they were more sensitive to the specification. For an assessment of different classes of models, including model averaging and neural networks, see the contemporaneous work of Guth (2022). Our agnostic approach also helps mitigate the concerns about model uncertainty when performing stress tests (Gross and Población, 2019).

[12] Appendix A provides more details about the definitions of all potential explanatory variables and the data sources. We use either the growth rate in log difference for variables expressed in dollars or the level directly for ratios or rates.

[13] In addition to the sources singled out in the main text, we also used the following: Makri, Tsagkanos, and Bellas (2014); Godlewski (2005); García-Marco and Robles-Fernández (2008); Louzis, Vouldis, and Metaxas (2012); Radivojevic and Jovovic (2017); Salas and Saurina (2002); Espinoza and Prasad (2010); Babouček and Jančar (2005); Rinaldi and Sanchis-Arellano (2006); Nkusu (2011); Mileris (2012); Figlewski, Frydman, and Liang (2012); Jovovic (2014); Škarica (2014); Beck, Jakubik, and Piloiu (2015); Rajan and Dhal (2003); Hoggarth, Logan, and Zicchino (2005); Jiménez and Saurina (2006); Saba, Kouser, and Azeem (2012); De Bock and Demyanets (2012); Pouvelle (2012); Shu (2002).

service issues and may foreshadow higher PDs (Beck et al., 2015; Nkusu, 2011).

Monetary policy and interest rate factors also have an impact on PDs, particularly those that represent the cost of borrowing because they have a direct impact on borrowers' debt-servicing capacity (e.g., five-year Government of Canada (GoC) bonds, six-month treasury bills, term premium, corporate spread). Inflation, in contrast, has a more ambiguous impact. While inflation can make servicing local currency debt easier by lowering its real value, it can also lead to higher nominal and real interest rates, raising debt servicing costs.

Moreover, the level of debt held by firms and other economic agents is a potentially important explanatory factor. High credit growth could indicate a credit boom with poor credit quality, resulting in higher future PDs (Schularick and Taylor, 2012; Danielsson, Valenzuela, and Zer, 2018; Kirti, 2018). In contrast, during a crisis, credit expansion can provide the liquidity required to keep firms from defaulting.

Aggregate financial conditions in the banking sector, particularly those related to banks' profitability and cost-efficiency (e.g., net income, return on equity) and assets and liabilities (e.g., total assets, total-assets-to-capital ratio, equity), could also be drivers of PDs (Podpiera and Weill, 2008) but have ambiguous relationships (Berger and DeYoung, 1997).[14]

Finally, total trade between Canada and the United States, the drop in oil prices from its two-year peak and the nominal USD/CAD exchange rate are used to capture foreign trade conditions. A decrease in total trade may lead to an increase in PDs due to lower firm revenues and profits. The relationship between the exchange rate and PDs is more ambiguous. On the one hand, it can cushion the decline in economic activity during times of stress, thereby aiding in the stabilization and reduction of

[14]Note that we do not estimate bank-specific models and focus instead on aggregate macrofinancial variables only. Thus, if bank-specific characteristics matter, they would not be captured here. Differences in the sector composition of the loan books across banks are captured afterward since we decompose aggregate default probabilities into sectoral components. For instance, the PD of a bank b could be computed as follows, given the bank-specific exposure to each industry $i \in ind$ and the industry-specific $PD_{i,t}$: $PD_{b,t} = \sum_{i \in ind} w_{i,b,t} \cdot PD_{i,t}$ with $w_{i,b,t} = \frac{L_{i,b,t}}{\sum_{j \in ind} L_{j,b,t}}$ and $L_{i,b,t}$ referring to the outstanding Canadian-dollar amount of loans to industry i by bank b.

PDs. A depreciation, on the other hand, may reduce borrowers' ability to repay foreign currency–denominated debt.

We also consider foreign equivalents of some of the domestic factors, i.e., US industrial production, the US financial stress index and the US term premium.

Table 1: Macrofinancial predictors considered with expected signs

Canadian factors
Economic conditions: real GDP (-), unemployment rate (+)
Credit aggregates: total private credit-to-GDP ratio (?), business credit-to-GDP ratio (?), real business credit (?), residential mortgage to household disposable income ratio (?), consumer credit to household disposable income ratio (?)
Trade: Total trade (-), USD/CAD rate (?)
Asset prices and financial conditions: corporate spread (+), TSX index (-), Real house prices (-), financial stress index (+)
Monetary policy and interest rates: inflation rate (?), 6-month t-Bills (+), 5-year GoC bonds (+), term premium (+), CA-US MP rate (+)
Banking: total assets (+), total assets-to-equity ratio (+), shareholders' equity (-), return on equity (-), net income (-)
US (foreign) factors
Economic conditions: industrial production (-)
Commodity prices: drop in oil prices (+)
Asset prices and financial conditions: financial stress index (+)
Monetary policy and interest rates: term premium (+)

Note: When the literature has reached a consensus on a sign, it is indicated in parentheses. When the expected sign is not well defined, a question mark is provided, as several different transmission channels leading to opposite signs have been identified in the literature.

3.2 How do we select the predictors?

Given the relatively small number of observations (from the first quarter of 1997 to the fourth quarter of 2019),[15] the limited number of historic financial crises, the large number of potential explanatory variables and the potential relevance of different lag orders, we face the risk of in-sample overfitting with poor out-of-sample performance.

[15] We have data starting a few quarters earlier, but they display a decreasing trend for some sectors that were recovering from the recession in the early 1990s. Including only part of the recovery phase of this business cycle would blur the relationship we aim to estimate.

We overcome this problem by using elastic net regularization (Zou and Hastie, 2005) as a model selection method, as well as expert judgement, to identify the most informative combination of PD predictors.

Elastic net regularization reduces the number of predictors by penalizing parameter fitting. It is a state-of-the art estimation procedure that extends the least squared objective by a penalization term based on the coefficient size. In other words, the procedure prevents overfitting by lowering the absolute value of the coefficients, partially setting them to zero, and thus including model selection.[16]

We select the model based on three criteria. First, based on the mean squared error (MSE), the selected model must either minimize the out-of-sample MSE (Min-MSE) or come within one standard deviation of it (1se-MSE). The latter is a conventional method for selecting predictors in a conservative manner. The 10-fold cross-validation technique is used to compute the out-of-sample MSEs.[17] The model's output must then be robust in terms of sign and variable selection for the majority of the $\lambda = \{1, 2, 3, 4\}$, the free parameter calibrated as part of the PD filtration process. Furthermore, the coefficients' signs must be as expected, or they can be rationalized.

Table 2 displays the Min-MSE and 1se-MSE for an autoregressive model of order up to 2 for various combinations of the regressors shown in Table 1: (1) current values; (2) current and first lagged values; and (3) current, first and second lagged values. Several results stand out. First, most MSE specifications provide comparable out-of-sample predictions. Second, because the Min-MSE specifications yield only a slight gain of about 2% on average when compared with the 1se-MSE, we take a conservative approach in terms of the number of predictors chosen. Third, while the specification with up to two lagged values has a better MSE on average, it has fewer common explanatory variables across values of λ, and some signs are inconsistent with

[16] The penalty parameter is a linear combination of the L1 and L2 penalties of the lasso (Tibshirani, 1996) and ridge (Hoerl and Kennard, 1970a, 1970b) methods.

[17] This method randomly divides the dataset into 10 different subsets. One of the subsets is kept as the validation set, and the model is trained (estimated) on the remaining 9 sets. We repeat this process using each of the 10 subsets as a validation set, resulting in 10 estimates of the MSE for each parameter value. The reported estimation is simply the average value of the 10 estimates.

expectations.

Table 2: Elastic net—MSEs

Criterion	X included	$\text{PD}^{\lambda=1}$	$\text{PD}^{\lambda=2}$	$\text{PD}^{\lambda=3}$	$\text{PD}^{\lambda=4}$
Min-MSE	t	0.1640	0.1183	0.0928	0.0738
Min-MSE	t, t-1	0.1652	0.1187	0.0923	0.0720
Min-MSE	t, t-1,t-2	0.1626	0.1130	0.0886	0.0722
1se-MSE	t	0.1663	0.1202	0.0943	0.0747
1se-MSE	t, t-1	0.1679	0.1208	0.0937	0.0729
1se-MSE	t, t-1, t-2	0.1663	0.1146	0.0897	0.0730

Note: Elastic net with 10-fold cross-validation and a maximum number of non-zero coefficients of 15. λ is the Poisson parameter explained in section 2.

Overall, we select the specification of 1se-MSE for parsimony with up to one lag of macrofinancial explanatory variables. Table 3 reports the full results for the selected specification for all $\lambda = \{1, 2, 3, 4\}$. According to this specification, the PDs are best explained by:

1. an AR(1) process
2. the current values of the unemployment rate, the real business credit, the policy rate differential between Canada and the United States, and the Canadian term premium
3. the first lagged values of the real business credit, the CAD/USD exchange rate, the five-year GoC bond yield, and the US financial stress index

Expert judgement is also used to guide the final specification. Because most stress test scenarios are designed around a GDP path, we do not penalize Canadian real GDP and US industrial production, and we include their lagged values. We also overrule the elastic net by including a second PD lag to remove serial correlation in the residuals. However, as a result of these modifications, the current Canadian term premium no longer bears the correct sign. Thus, we use its lagged value instead, rendering the US financial stress index obsolete (i.e., with a marginal impact on the out-of-sample performance). Table 4 displays the final list of variables.

Table 3: Elastic Net—coefficients of 1se-MSE with up to 1 lag

Variables	Transf.	$PD^{\lambda=1}$	$PD^{\lambda=2}$	$PD^{\lambda=3}$	$PD^{\lambda=4}$
PD^{λ}_{t-1}	level	0.8416	0.8431	0.8347	0.8368
PD^{λ}_{t-2}	level	0	0	0	0
Real GDP_t	g.r.	0	0	0	0
Real GDP_{t-1}	g.r.	0	0	0	0
Unemployment $rate_t$	level	0.0001	0.0098	0.0108	0
Unemployment $rate_{t-1}$	level	0	0	0	0.0071
Total private credit-to-GDP $ratio_t$	g.r.	0	0	0	0
Total private credit-to-GDP $ratio_{t-1}$	g.r.	0	0	0	0
Business credit-to-GDP $ratio_t$	g.r.	0	0	0	0
Business credit-to-GDP $ratio_{t-1}$	g.r.	0	0	0	0
Real business $credit_t$	g.r.	-0.0005	-0.0007	-0.0001	0
Real business $credit_{t-1}$	g.r.	0	-0.0029	-0.0070	-0.0069
Total $trade_t$	g.r.	0	0	0	0
Total $trade_{t-1}$	g.r.	0	0	0	0
USD/CAD $rate_t$	level	0	0	0	0
USD/CAD $rate_{t-1}$	level	-0.2539	-0.1573	-0.1234	-0.0580
Corporate $spread_t$	level	0	0	0	0
Corporate $spread_{t-1}$	level	0	0	0	0
TSX $index_t$	g.r.	0	0	0	0
TSX $index_{t-1}$	g.r.	0	0	0	0
Real house $prices_t$	g.r.	0	0	0.0005	0.0020
Real house $prices_{t-1}$	g.r.	0	0	0	0.0010
Financial stress $index_t$	level	0	0	0	0
Financial stress $index_{t-1}$	level	0	0	-0.0399	-0.1390
Inflation $rate_t$	level	0	0	0	0
Inflation $rate_{t-1}$	level	0	0	0	0
6-month $t\text{-}Bills_t$	level	0	0	0	0
6-month $t\text{-}Bills_{t-1}$	level	0	0	0	0
5-year GoC $bonds_t$	level	0	0	0	0
5-year GoC $bonds_{t-1}$	level	0.0051	0.0020	0.0059	0.0068
Term $premium_t$	level	0.0691	0.0662	0.0603	0.0410
Term $premium_{t-1}$	level	0	0	0.0110	0.0229
CA-US MP $rate_t$	level	0.0121	0.0093	0.0235	0.0300
CA-US MP $rate_{t-1}$	level	0	0	0	0
Total $assets_t$	g.r.	0	0	0	0
Total $assets_{t-1}$	g.r.	0	0	0	0
Total assets-to-equity $ratio_t$	level	0	0	0	0
Total assets-to-equity $ratio_{t-1}$	level	0	0	0	0
Shareholders' $equity_t$	g.r.	0	0	0	0
Shareholders' $equity_{t-1}$	g.r.	0	0	-0.0008	-0.0015
Return on $equity_t$	level	0	0	0	0
Return on $equity_{t-1}$	level	-1.0186	-0.2569	0	0
Net $income_t$	g.r.	0	0	0	0
Net $income_{t-1}$	g.r.	0	0	0	0
US industrial $production_t$	g.r.	0	0	0	0
US industrial $production_{t-1}$	g.r.	0	0	0	0
Oil $prices_t$	g.r.	0	0	0	0
Oil $prices_{t-1}$	g.r.	0	0	0	0
US financial stress $index_t$	level	0	0	0	0.0253
US financial stress $index_{t-1}$	level	0.2251	0.1454	0.0865	0.0772
US term $premium_t$	level	0	0	0	0
US term $premium_{t-1}$	level	0	0	0	0

Note: Elastic net with 10-fold cross-validation and a maximum number of nonzero coefficients of 15. λ is the Poisson parameter explained in section 2.

3.3 Linear model: Autoregressive model

The final linear model, an autoregressive model of order 2 with additional predictors in the vector X (henceforth ARX), takes the following form:

$$PD_t^{\lambda} = \alpha + \sum_{i=1}^{2} \beta_i PD_{t-i}^{\lambda} + \gamma X_t + \epsilon_t. \tag{6}$$

Table 4 shows our estimates for equation (6) for PD_t^{λ} with the free parameter for the PD filtration $\lambda = \{1, 2, 3, 4\}$. The majority of our estimates are consistent with our beliefs. First, the persistence of PD_t^{λ} is high, with estimates ranging from 0.87 for $\lambda = 1$ and 0.93 for $\lambda = 4$.[18] Second, the coefficient for the unemployment rate is positive, whereas those for real GDP and US industrial production are negative, indicating that an improvement in economic conditions reduces PD_t^{λ}. Next, the interest rates (i.e., five-year GoC bonds and US term premium) are positive, indicating that PD_t^{λ} will rise as the cost of borrowing rises. Finally, tighter credit conditions, as measured by a decrease in real business credit, increase PD_t^{λ}, implying that access to additional credit lines is advantageous and outweighs, at least in the short term to medium term, the financial vulnerability channel represented by increased business indebtedness.

The autoregressive coefficients are statistically significant, while most other variables are not. The p-values report the level of marginal significance, which is based on an in-sample goodness-of-fit measure. However, in the context of a stress test, the objective of our model is to produce out-of-sample prediction. Table 5 compares the performance of our model with that of a simple AR(2) model to assess the value added by the additional predictors. It presents the average root mean squared forecast error (RMSE) of rolling 12-quarter-ahead forecasts made over the period $t = 10$ through the end of the sample. The first row contains the full ARX model, which takes the exogenous regressors over the projection horizon as given; and the second row contains

[18] The persistence is the sum of the coefficients on the lagged endogenous variables.

Table 4: Estimated parameters—ARX

Variables	Transf.	$PD^{\lambda=1}$	$PD^{\lambda=2}$	$PD^{\lambda=3}$	$PD^{\lambda=4}$
PD^{λ}_{t-1}	level	1.4235***	1.5542***	1.5832***	1.5868***
PD^{λ}_{t-2}	level	-0.5585***	-0.6567***	-0.6687***	-0.6598***
Unemployment rate$_t$	level	0.0107	0.0095	0.0092	0.0081***
Real business credit$_t$	g.r.	-0.0117	-0.0153**	-0.0132**	-0.0098
CA-US MP rate$_t$	level	0.0452	0.0185	0.0090	0.0044
Real GDP$_{t-1}$	g.r.	-0.0227	-0.0198	-0.0116	-0.0095
Real business credit$_{t-1}$	g.r.	-0.0049	-0.0035	-0.0056	-0.0058***
USD/CAD rate$_{t-1}$	level	-0.2525*	-0.0939	-0.0442	-0.0268
5-year GoC bonds$_{t-1}$	level	0.0230**	0.0135**	0.0090	0.0060***
Term premium$_{t-1}$	level	0.0591*	0.0308	0.0209	0.0182
US industrial production$_{t-1}$	g.r.	-0.0129	-0.0080	-0.0080	-0.0063

Note: Results from estimation of equation (6). All models include an intercept. The symbols *, ** and *** indicate statistical significance of the coefficient at the 10%, 5% and 1% levels, respectively. λ is the Poisson parameter explained in section 2.

the simple AR(2) model. The ARX model outperforms in this projection, with a gain of over 60% in the last quarter.

Table 5: h-quarter-ahead forecast RMSE—ARX versus AR(2)

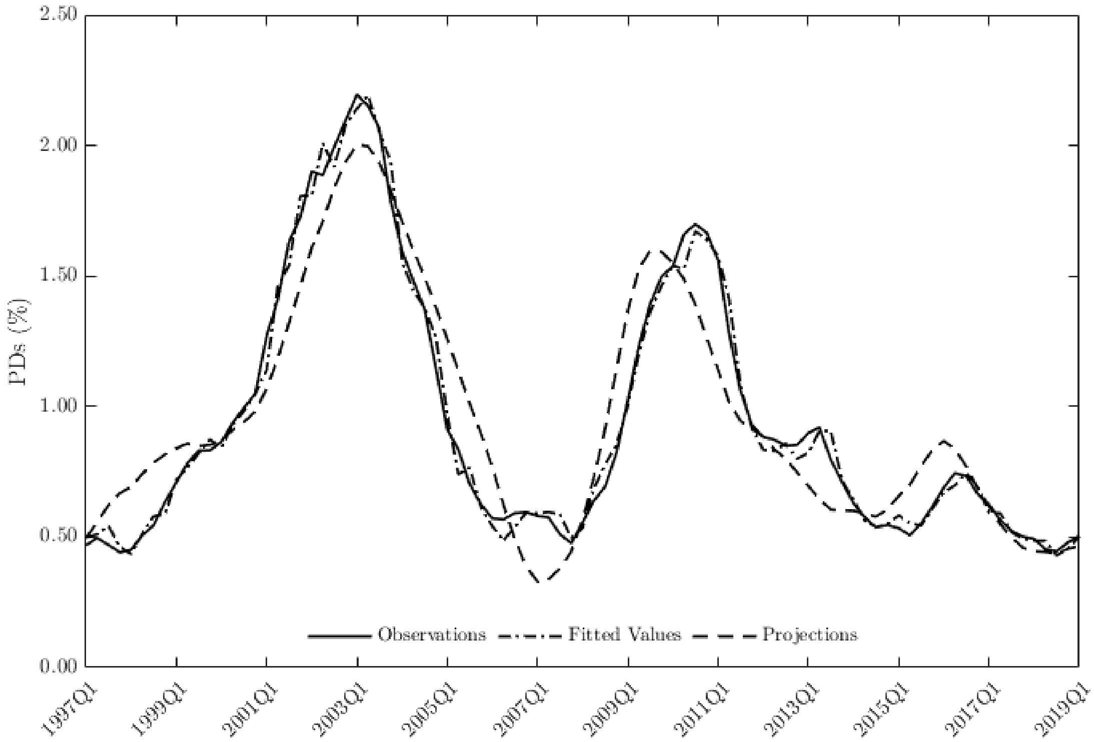
h-quarter ahead	1	2	3	4	5	6	7	8	9	10	11	12
ARX	0.06	0.10	0.15	0.17	0.19	0.20	0.21	0.20	0.20	0.19	0.19	0.18
AR(2)	0.07	0.13	0.20	0.25	0.31	0.36	0.40	0.43	0.46	0.47	0.48	0.48

Note: Results from simulations of equation (6) for $\lambda = 3$. Results are similar for $\lambda = \{1, 2, 4\}$.

Figure 4 provides a visualization of the performance of the model by presenting its predictions with the fitted values, as well as projections of PDs conditional only on the observed predictors in X.[19] The figure shows that the full ARX model has an excellent in-sample fit. The ARX model fits the levels and timing of the swings in PDs better than the simpler AR(2) model.

[19]Except for the first two periods, which are used to initialize the projections, we use the projected values of PDs of the previous periods as lagged explanatory variables for the next period, not the observed one. Therefore, when we roll the projections forward, the projection errors are accumulated over time.

Figure 4: Model fit—ARX



Note: Results from estimation of equation (6) for $\lambda = 3$. Results are similar for $\lambda = \{1, 2, 4\}$.

3.4 Non-linear model: Quantile Regression Model

Although the autoregressive model predicts well, a linear specification is restrictive. It implies, among other things, that shock responses will be symmetric (i.e., the impact of positive and negative shocks of the same magnitude will be the same in absolute terms) and independent of history (i.e., the impact does not depend on the starting point).

To overcome these constraints, we model the non-linear and non-normal responses of PDs using conditional quantile regressions, an extension of standard linear regressions. This is especially helpful in understanding outcomes that are asymmetrically distributed and have non-linear relationships with predictor variables. It allows us to relax the assumption that the dynamics of PDs are the same at the distribution's upper and lower tails. We can also determine which factors are most important in causing the large increase in PDs, which is useful for stress testing.

Equation (6) is reformulated into a quantile regression (QR) framework. Formally, the QR estimators minimize:

$$Q(\beta_q) = \sum_{t:PD_t^\lambda \geq \widehat{PD_t^\lambda}} q|PD_t^\lambda - \widehat{PD_t^\lambda}| + \sum_{t:PD_t^\lambda < \widehat{PD_t^\lambda}} (1-q)|PD_t^\lambda - \widehat{PD_t^\lambda}|, \tag{7}$$

where $\widehat{PD_t^\lambda} = \alpha_q + \sum_{i=1}^{2} \beta_{i,q} PD_{t-i}^\lambda + \gamma_q X_t$ is the fitted value for quantile q, which takes values from 0.1 to 0.9. Equation (7) is estimated using the majorize-minimize method of Hunter and Lange (2000).[20]

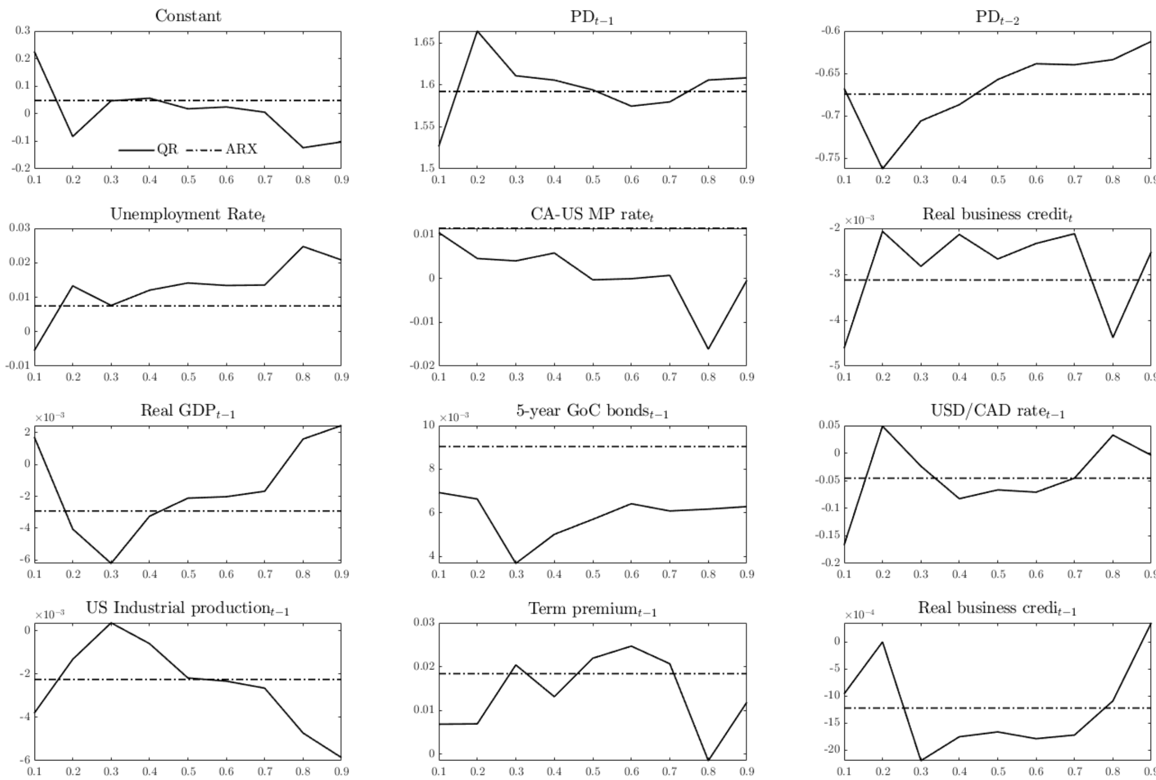
In Figure 5, we compare the estimated parameters of the autoregressive parameters (represented by the horizontal dashed-dotted black lines) with those of quantile regression (represented by the solid black lines), which differ by quantile from 0.1 to 0.9.

We find these dynamics particularly interesting. First, the autoregressive process is more persistent in the upper tail than in the lower tail, implying that PDs rise faster and are more persistent during a crisis than in normal times. Furthermore, the unemployment rate and US industrial production have an impact that is two to six times greater on PDs in the upper quantiles than in the lower quantiles. On the other hand, the effects of real GDP and real business credit seem muted in the upper tails. Given that the model will be used in the context of stress testing, with scenarios that are frequently in the upper tail of the distribution, these accentuated reactions in the upper quantiles are better suited to simulations than to a linear model.

Despite the fact that the differences in estimated parameters between quantiles appear to be minor, they are statistically significant. The test statistics and p-values of a Wald test for equality of slope coefficients between quantiles with the null hypothesis of no significant difference are shown in Table 6. The p-values show that the upper quantile differs significantly from the lower quantiles.

[20] The results are similar using the interior point algorithm of Koenker and Park (1996) for the quantile regression.

Figure 5: Estimated parameters—QR versus ARX models



Note: Results from estimation of equation (6) and equation (7) with $\lambda = 3$. Results are similar for $\lambda = \{1, 2, 4\}$.

Table 6: Quantile slope equality test

Tests	Q95–Q50	Q90–Q50	Q90–Q10	Q95–Q05	Q50–Q10	Q50–Q05
χ^2 statistics	7.18	9.78	17.51	22.96	17.27	19.00
p-values	0.78	0.55	0.09	0.02	0.10	0.06

Note: Wald test of equality of the slope parameters between two quantiles.

3.5 Backtesting historical PDs

We now provide a backtesting of the models in equations (6) and (7) to evaluate how they would have performed if they had been available at different times, particularly during stressful periods. The goals are to assess the models' viability and gain confidence in their use in the future, as well as to help calibrate the future stress test scenario analysis to specific quantiles capable of replicating past crisis periods.

Figure 6 depicts the ability of the models to replicate the level of PDs observed during the two main peaks. Panel a of the figure depicts the dynamics of the ARX model, whereas panel b depicts the various dynamics that can be generated using the quantile models. The figure shows that the ARX model performs well, with the upper confidence interval reproducing the historical PD peaks. However, simulating a severe crisis for stress-testing purposes with a linear model based on mild crises may underestimate the increase in PDs.

In contrast, the dynamics implied by the quantile models can generate even larger PDs during the crisis. This is due to the accentuated responses in the tail of the distribution, and the fact that we iterate the simulation forward using the values projected for the quantile of the PD at the previous period. For instance, the simulation of the 90^{th} percentile in panel b assumes that the dynamics of the PD remain those of the 90^{th} percentile throughout the whole simulation—that is, it uses the elasticities estimated with the quantile regression of the 90^{th} percentile of the distribution. In a simulation, if we keep for an extended period of time the dynamics of the PD to be those of a crisis period, we can then simulate extreme PD paths previously unobserved. This is especially important given that Canada has not experienced a severe banking crisis in recent history.

The dashed line in panel b shows a combination of quantiles from the QR model, starting the simulation at the 70^{th} percentile and then gradually moving to the 20^{th} percentile (a proxy for a tranquil period) of the PD distribution. This simulation is able to reproduce historical stress episodes. It suggests that, for stress-testing purposes, when designing scenarios we should use at least the 70^{th} quantile of the QR-ARX model.

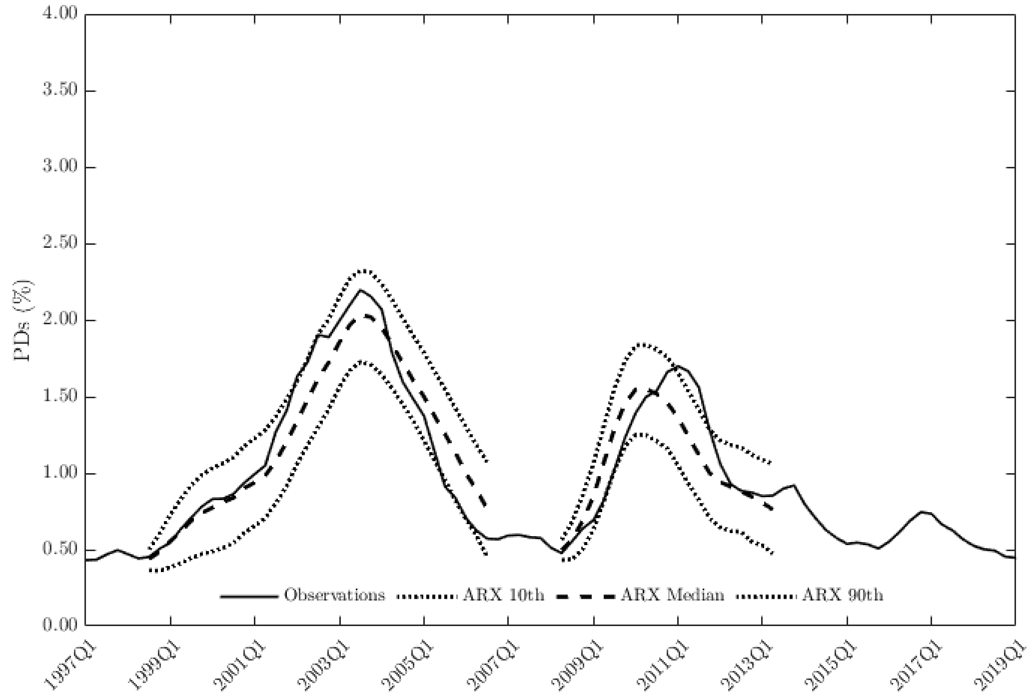
4 Modelling industry probability of default

In this section, we model PDs for each of the 10 industries. This modelling strategy proceeds in a fashion similar to the aggregate presented in section 3, with three variations.

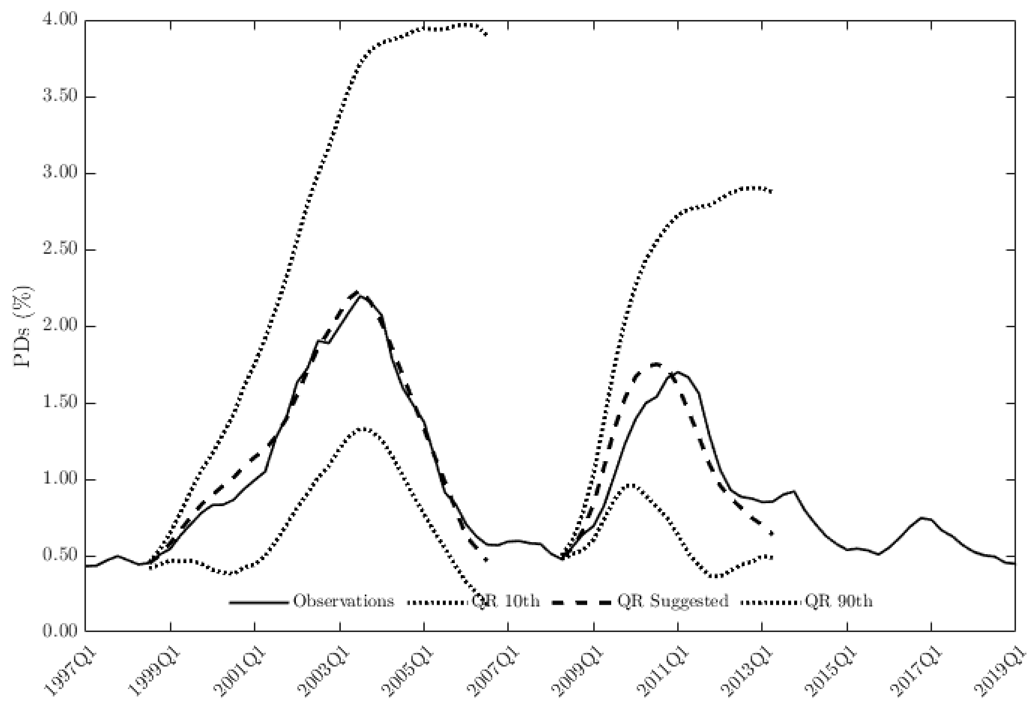
First, we perform variable selection for each industry using the elastic net regularization procedure described in section 3. However, the set of potential predictors is

Figure 6: Backtesting—ARX and QR

(a)



(b)



limited to those chosen for the aggregate PD, to which we add a few variables that may have an industry-specific impact (e.g., house prices for construction and real estate; the drop in oil prices from their maximal values over two years for mining, quarrying and oil).

Second, we also include two aggregate PD lags as potential predictors to anchor the dynamic of the industry-specific PDs around the aggregate.

Finally, we focus on a linear autoregressive model for industries, because we account for non-linear dynamics by including lagged aggregate PDs from the quantile regressions.

The final linear model, an autoregressive model of order 2 with additional predictors (henceforth Ind-ARX), takes the following form:

$$PD_t^{\lambda,ind} = \alpha + \sum_{i=1}^{2} \beta_i^{ind} PD_{t-i}^{\lambda,ind} + \sum_{i=1}^{2} \beta_i PD_{t-i}^{\lambda} + \gamma X_t + \epsilon_t, \tag{8}$$

where $PD^{\lambda,ind}$ is the industry-specific PD and PD^{λ} is the aggregate PD. The final selection of predictors, along the estimated parameters, is presented in Table 7 for each industry for PD_t^{λ} with $\lambda = 3$.[21]

Several outcomes should be highlighted. First, the persistence of industry-specific PDs is heterogeneous, with the majority being lower than the persistence of the aggregate PD process. This is most likely due to the inclusion of the latter as an anchor that captures a portion of the persistence.[22] Second, for most industries, an improvement in economic conditions reduces PDs, an increase in borrowing costs increases PDs, and a credit tightening through a decrease in real business credit raises PDs. Third, some of those patterns differ, especially for the mining, quarrying and oil industries. Higher domestic borrowing costs are associated with lower PDs, but this is likely because international funding conditions and the price of oil in US dollars are what matter most, as reflected by the significant effect of the USD/CAD exchange rate.

[21] Results are similar for $\lambda = \{1, 2, 4\}$.

[22] Lags of the industry-specific PDs are always selected when included as potential predictors in the elastic net regularization procedure for each industry.

Table 7: Estimated parameters—industry-specific ARX

Variables	Natural Resources	Mining, Quarrying, and Oil	Manufacturing	Construction and Real Estate	Transportation, Communications, and Utilities
$\text{PD}_{t-1}^{\lambda=3,ind}$	1.0369***	1.5404***	1.3762***	1.7783***	1.4171***
$\text{PD}_{t-2}^{\lambda=3,ind}$	-0.2964***	-0.6674***	-0.4399***	-0.8577***	-0.5441***
$\text{PD}_{t-1}^{\lambda=3}$	0.0409	0.2210	0.1984	0.0256	0.1956
$\text{PD}_{t-2}^{\lambda=3}$	0.0449	-0.0796	-0.2571	-0.0310	0.0803
Unemployment rate$_t$	0.0650***		0.0103	0.0469**	
Real business credit$_t$	-0.0104			-0.0130	-0.0250
Real house prices$_t$				-0.0033	
CA-US MP rate$_t$		-0.0472	0.0060	-0.0189	0.1205*
Real GDP$_{t-1}$				-0.0262	-0.0823
Real business credit$_{t-1}$	-0.0142	-0.0731***	0.0090	-0.0143*	-0.0275
USD/CAD rate$_{t-1}$		-0.7466**	0.0390	0.3542**	-0.9079**
5-year GoC bonds$_{t-1}$	0.0300***	-0.0611**	0.0145	-0.0025	0.0395
Term premium$_{t-1}$			0.0590*	0.0045	
US Industrial production$_{t-1}$			-0.0196*		
Oil price drop$_{t-2}$		0.2664			
Variables	Wholesale Trade	Retail	Service	Other Corporates	Financial Institutions
$\text{PD}_{t-1}^{\lambda=3,ind}$	1.2206***	1.1094***	1.3582***	1.4898***	1.2206***
$\text{PD}_{t-2}^{\lambda=3,ind}$	-0.2987***	-0.3152***	-0.4844***	-0.6012***	-0.3556***
$\text{PD}_{t-1}^{\lambda=3}$	0.3390***	-0.0022	0.1423	0.0269	0.0180*
$\text{PD}_{t-2}^{\lambda=3}$	-0.3545***	0.0087	-0.1139	-0.0680	-0.0196**
Unemployment rate$_t$	0.0101	0.0685***	0.0162	0.0002	0.0038***
Real business credit$_t$	0.0062	0.0088	-0.0071	-0.0356***	
Real house prices$_t$					
CA-US MP rate$_t$			-0.0129		
Real GDP$_{t-1}$			-0.0096	-0.0061	
Real business credit$_{t-1}$	-0.0034	0.0132	-0.0054	-0.0058	
USD/CAD rate$_{t-1}$			0.3025**		
5-year GoC bonds$_{t-1}$	0.0007	0.0235***		0.0222***	0.0001
Term premium$_{t-1}$	0.0113	0.0331	0.0156	0.0748***	
US Industrial production$_{t-1}$	-0.0102				
Oil price drop$_{t-2}$					

Note: Results from estimation of equation (8) for $\lambda = 3$. Results are similar for $\lambda = \{1, 2, 4\}$. All models include an intercept. The symbols *, ** and *** indicate statistical significance of the coefficient at the 10%, 5% and 1% levels, respectively.

As expected, a persistent drop in oil prices compared with the peak reached over the previous two years increases PDs in this sector. However the increase is not significant, likely because of the importance of the oil sector for the Canadian economy: when oil prices are low, that already tends to depreciate the Canadian dollar, which is already reflected in the USD/CAD exchange rate.

Figure 7 depicts the performance of the industry-specific models by displaying fitted values as well as projections conditional solely on the observed external predictors

X.[23] The figure shows that the majority of the industry-specific ARX models have a good in-sample fit and can fit both the levels and the timing of the swings in industry PDs.

The autoregressive coefficients, like the aggregate PD model, are statistically significant, whereas the majority of the other variables are not. Table 8 assesses the value added of the two aggregate PD lags and the exogenous variables as predictors compared with a simple AR(2) model. It displays the industry-specific average RMSE of rolling 12-quarter-ahead forecasts made from the first quarter of 1999 to the fourth quarter of 2019. The use of a 12-quarter projection horizon is consistent with what is typically used in stress-testing exercises. The first row shows the RMSE for the industry-specific ARX models. The second row shows the RMSE of the industry-specific AR(2) models that include the two aggregate PD lags but exclude the exogenous regressors (denoted AR(2)+PD$^{\lambda}$). The third row shows the RMSE of the simple industry-specific AR(2) models that exclude the aggregate PD lags and the exogenous regressors (denoted AR(2)). We find that the Ind-ARX model improves upon the AR(2)+PD$^{\lambda}$ model for all industries, and, in turn, the AR(2)+PD$^{\lambda}$ model improves upon the simple AR(2) models. However, the difference is less significant for the financial institutions, owing to the lower level of the PD in this industry and less volatility to explain over time.

Aggregate industry-specific PDs. While the aggregate PD is used as an anchor in each industry-specific model, there are no constraints in the estimation that ensure the industry-specific PD projections are consistent with the aggregate PD projection for the same stress-test scenario.

To address this inconsistency, we first compute a weighted average of industry-specific PDs:

$$PD_{T+h}^{\lambda,wa} = \sum_{ind} PD_{T+h}^{\lambda,ind} \frac{TL_T^{ind}}{TL_T},$$

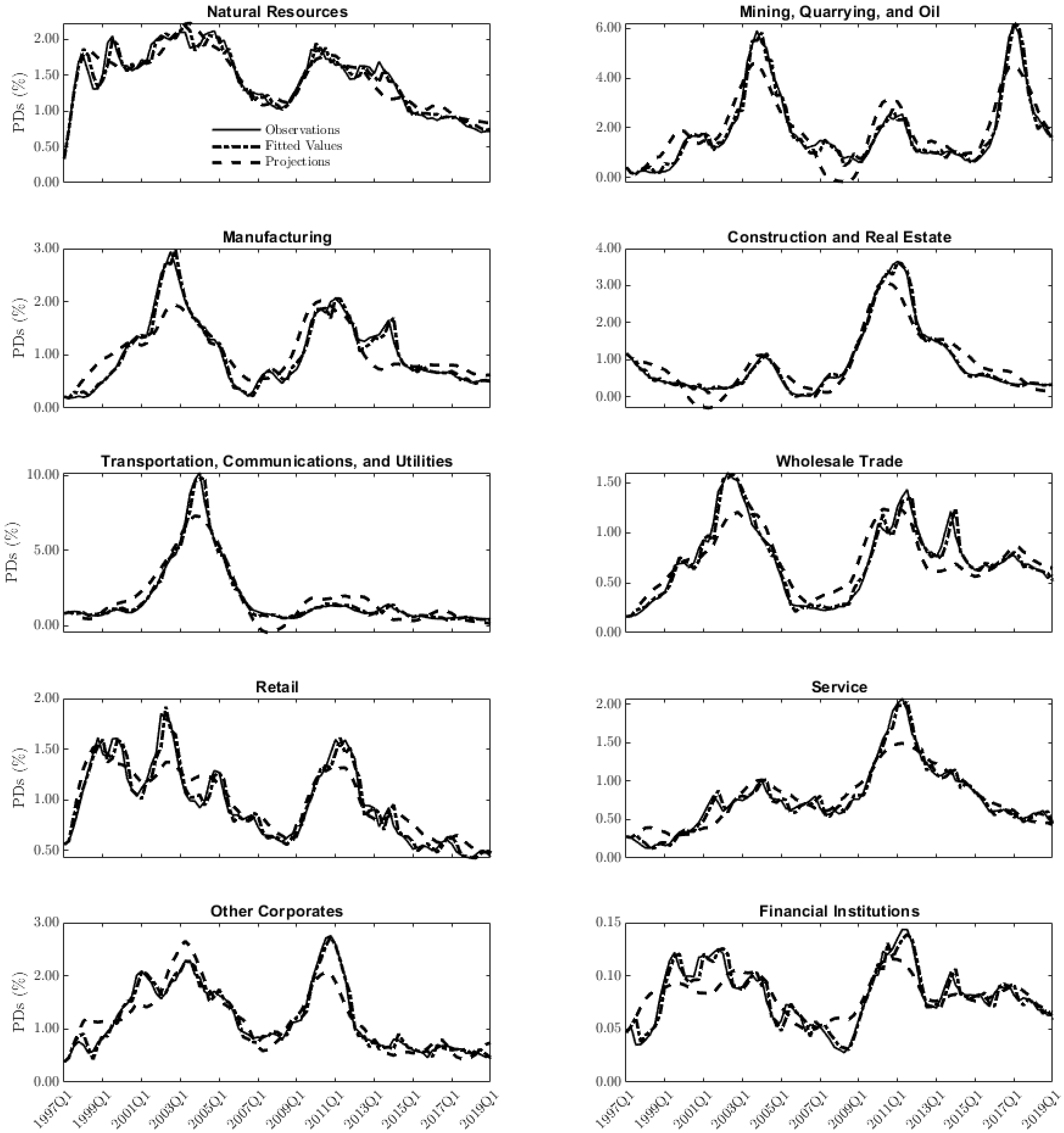
[23] The projections are generated the same way as for the aggregate PD projections of Figure 4, with projection errors accumulating progressively as the time horizon expands.

Table 8: h-quarter-ahead forecast RMSE—industry-specific ARX versus AR(2)+PD$^{\lambda}$ vs AR(2)

h-quarter ahead	1	2	3	4	5	6	7	8	9	10	11	12
Natural resources												
Ind-ARX	0.09	0.12	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
AR(2)+PD$^{\lambda}$	0.10	0.14	0.19	0.21	0.23	0.25	0.26	0.27	0.27	0.27	0.27	0.27
AR(2)	0.10	0.15	0.21	0.25	0.29	0.32	0.36	0.38	0.39	0.40	0.41	0.42
Mining, quarrying and oil												
Ind-ARX	0.25	0.42	0.56	0.66	0.70	0.71	0.71	0.71	0.71	0.71	0.72	0.72
AR(2)+PD$^{\lambda}$	0.29	0.56	0.81	1.03	1.19	1.29	1.33	1.35	1.34	1.32	1.31	1.32
AR(2)	0.29	0.56	0.84	1.08	1.27	1.40	1.47	1.50	1.49	1.47	1.47	1.49
Manufacturing												
Ind-ARX	0.12	0.21	0.27	0.31	0.35	0.37	0.38	0.38	0.37	0.36	0.35	0.33
AR(2)+PD$^{\lambda}$	0.13	0.23	0.31	0.36	0.40	0.42	0.44	0.44	0.44	0.43	0.41	0.39
AR(2)	0.13	0.25	0.35	0.43	0.51	0.55	0.58	0.60	0.61	0.61	0.60	0.58
Construction and real estate												
Ind-ARX	0.08	0.16	0.24	0.29	0.31	0.33	0.33	0.33	0.32	0.31	0.30	0.30
AR(2)+PD$^{\lambda}$	0.10	0.21	0.35	0.47	0.56	0.65	0.72	0.78	0.82	0.85	0.87	0.89
AR(2)	0.10	0.22	0.36	0.48	0.58	0.67	0.75	0.81	0.86	0.89	0.91	0.92
Transportation, communications and utilities												
Ind-ARX	0.30	0.52	0.71	0.78	0.79	0.78	0.77	0.76	0.75	0.74	0.74	0.74
AR(2)+PD$^{\lambda}$	0.34	0.63	0.93	1.11	1.23	1.30	1.35	1.39	1.41	1.42	1.42	1.42
AR(2)	0.35	0.70	1.09	1.37	1.60	1.77	1.93	2.06	2.17	2.25	2.31	2.34
Wholesale trade												
Ind-ARX	0.08	0.13	0.16	0.17	0.18	0.19	0.20	0.21	0.21	0.20	0.20	0.19
AR(2)+PD$^{\lambda}$	0.08	0.13	0.17	0.18	0.19	0.20	0.21	0.21	0.21	0.20	0.20	0.19
AR(2)	0.09	0.16	0.22	0.26	0.30	0.32	0.34	0.35	0.36	0.36	0.35	0.35
Retail												
Ind-ARX	0.09	0.14	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.16	0.15	0.14
AR(2)+PD$^{\lambda}$	0.10	0.17	0.22	0.25	0.28	0.29	0.29	0.29	0.29	0.29	0.28	0.27
AR(2)	0.10	0.18	0.24	0.28	0.33	0.35	0.36	0.36	0.37	0.37	0.36	0.35
Service												
Ind-ARX	0.07	0.12	0.15	0.16	0.17	0.18	0.18	0.18	0.17	0.16	0.16	0.16
AR(2)+PD$^{\lambda}$	0.08	0.14	0.18	0.22	0.25	0.28	0.29	0.31	0.32	0.33	0.33	0.34
AR(2)	0.08	0.14	0.20	0.24	0.29	0.32	0.35	0.37	0.38	0.39	0.40	0.41
Other corporates												
Ind-ARX	0.10	0.17	0.22	0.26	0.29	0.29	0.28	0.27	0.26	0.26	0.26	0.26
AR(2)+PD$^{\lambda}$	0.12	0.22	0.30	0.36	0.40	0.42	0.42	0.42	0.41	0.42	0.42	0.43
AR(2)	0.13	0.24	0.34	0.43	0.50	0.55	0.58	0.60	0.61	0.62	0.63	0.64
Financial institutions												
Ind-ARX	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
AR(2)+PD$^{\lambda}$	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
AR(2)	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03

Note: Results from estimation of equation (8) for $\lambda = 3$. Results are similar for $\lambda = \{1, 2, 4\}$.

Figure 7: Model fit by industry



Note: Results from estimation of equation (8) with $\lambda = 3$. Results are similar for $\lambda = \{1, 2, 4\}$.

where h is the quarter ahead in the projection and the weights are the share of loans of each industry *ind* over total loans for the most recently observed quarter, which are then assumed to be invariant for the period of the projection.

We then compare the aggregate PDs with the weighted average of projected industry-specific PDs:

$$PDR^{\lambda}_{T+h} = \frac{PD^{\lambda,wa}_{T+h}}{PD^{\lambda}_{T+h}}$$

For each period, if the weighted average of industry-specific PDs is below or above the aggregate PDs, then we respectively scale up or down the industry-specific PDs by PDR^{λ}_{T+h}. This adjustment ensures that the aggregation of industry-specific PDs matches the aggregate PD.

5 Application

We now provide an example of credit risk assessment under an extreme but plausible macroeconomic scenario using RAMM (Tuzcuoglu, forthcoming). We simulate a scenario from the first quarter of 2021 onward until the end of 2024. This fictitious scenario describes economic developments in Canada and around the world.

Macrofinancial scenario. The severity and persistence of our scenario is aligned with the 1981–82 recession, the worst recession experienced in Canada (Table 9). The simulated peak-to-trough contraction of real GDP is -5.4% over six quarters, which generates low inflation pressures. Unemployment reaches a peak of 13%, an increase of nearly 4.5 percentage points over the fourth-quarter 2020 level. This recession is broad-based and synchronized with a global recession as well as with heightened stress in residential and commercial real estate, commodity prices and corporate debt markets. Given the period of low rates at the beginning of the scenario compared with other recession episodes, the three-month treasury rate remains near zero with the five-year Government of Canada bond yield averaging 0.5% over the scenario period. Finally, we assume tighter borrowing conditions for corporates, resulting in a sharp deleveraging of businesses.

Assessment of credit risk. Figure 8 depicts the impact of this extremely severe but plausible macrofinancial scenario on banks' aggregate corporate probabilities of default. The projected PD for the linear ARX model is represented by a solid black

Table 9: Hypothetical risk scenario—overview and comparisons

	Scenario	1981-82 recession	1990-91 recession	2008-09 recession
Real GDP contraction (peak to trough, %)	−5.4	−5.4	−3.4	−4.5
Recession duration (# of consecutive quarters of negative growth)	6	6	4	3
Peak unemployment	13.0	13.0	11.7	8.6
5-year GoC bond yields	0.5	15.0	10.6	2.25
Real business credit contraction (peak to trough, %)	−3.5	−6.0	−5.0	−0.5

line, along with its 80% confidence bands represented by dotted black lines. The PDs for the non-linear QR models are represented by dashed black lines for the 10^{th} and 90^{th} percentiles. For this simulation, we use as a starting point the probability of default in the third and fourth quarters of 2020 as filtered by our new method.[24]

Two key aspects stand out. First, when using the linear model, the projected PD already increases significantly. After two years, the PD reaches about 1.6% before gradually declining. As expected, using the quantiles of the QR models yields a broader range of potential projections. We can project a higher and more persistent PD by using higher quantiles, as shown by the 90^{th} percentile.

Second, the projected PD reaches levels comparable to the worst historical episodes. It is driven by the small increase in PDs at the beginning of the simulation period via the autoregressive parameters, the high rate of unemployment in the scenario and the sharp decrease in the level of business credit.

[24] Other data sources can be used as starting points for projections as long as they represent a quarterly PD and contain the industry decomposition, such as RAPID2 or the newly published IFRS 9 reports.

Figure 8: Projections of aggregate PD—hypothetical scenario

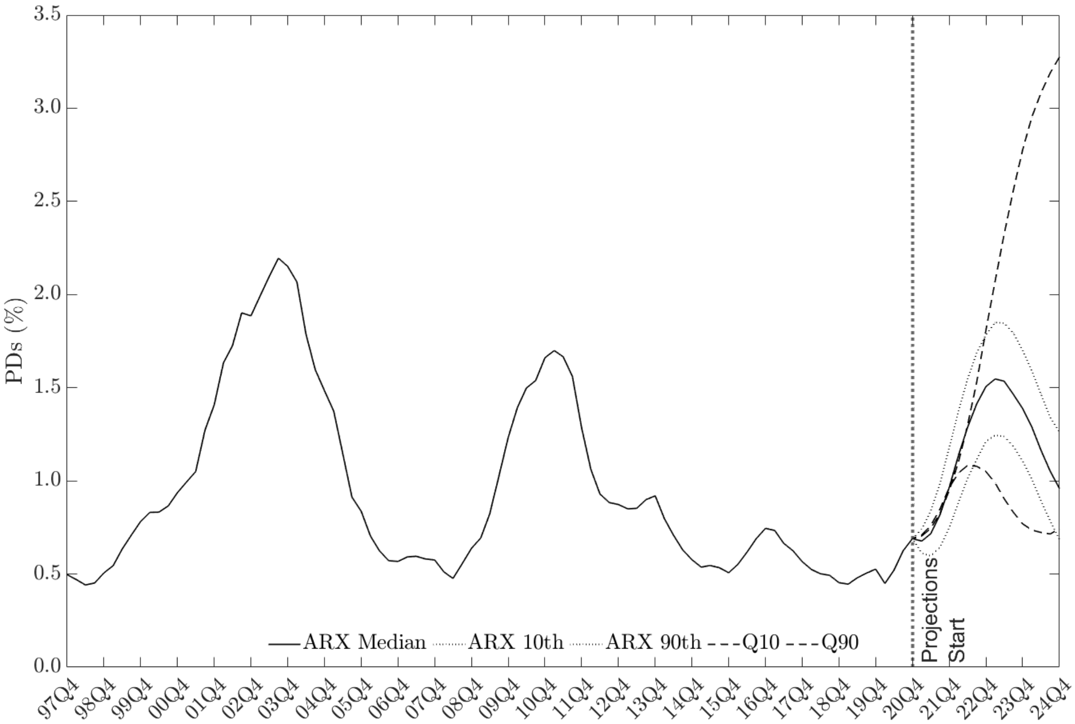
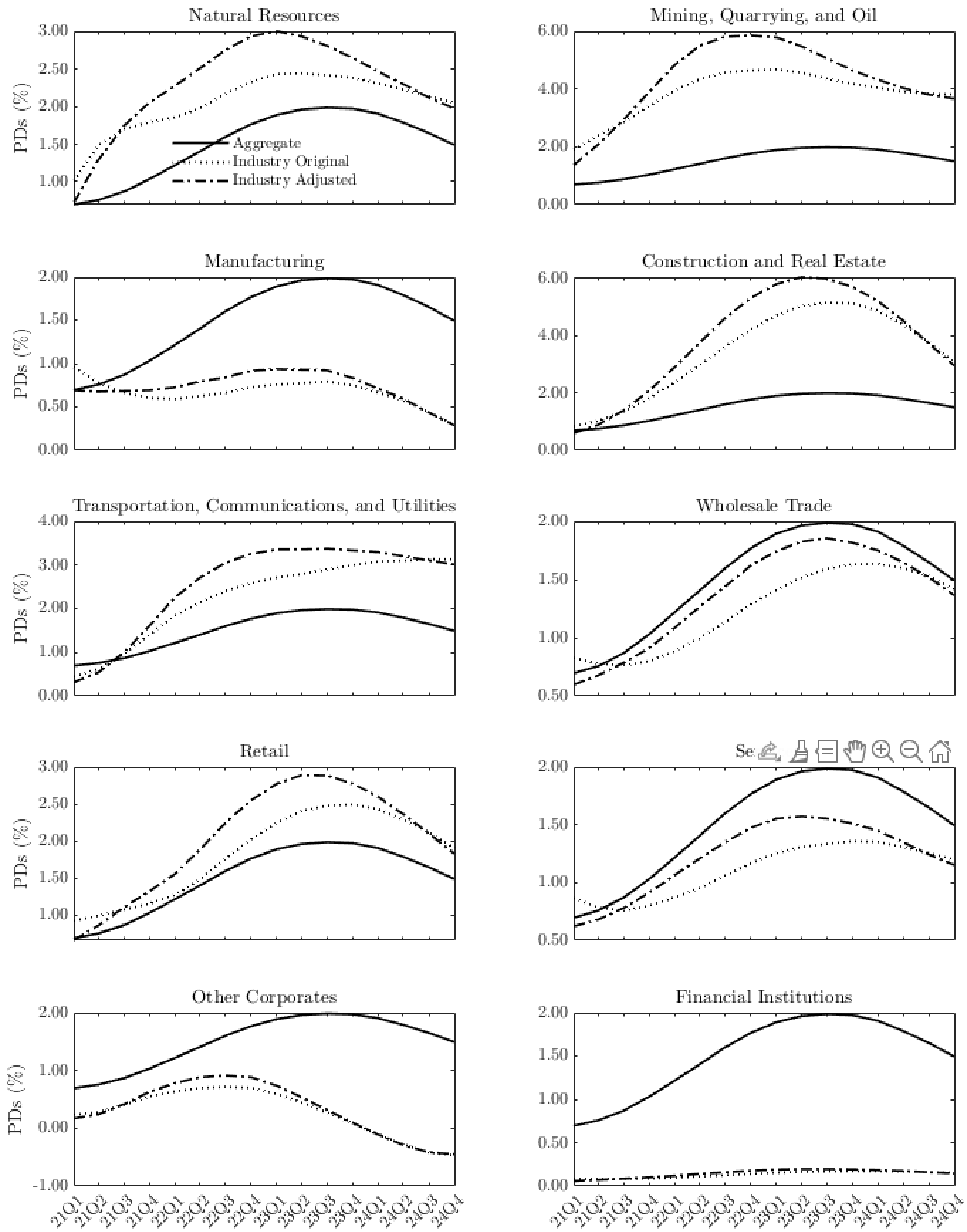


Figure 9 displays the projected PDs now split by industry. The line labelled "aggregate" is the aggregate PD, computed as a linear combination of the models of the 70^{th} and the 20^{th} quantiles as suggested in section 3.5. The same aggregate PD line is reproduced on each sub-figure for reference, along with the industry-specific PDs. The original industry-specific PD before adjustment is the dotted black line (i.e., the pure output of the industry model that may not perfectly aggregate up to the output of the aggregate model). The adjusted industry-specific PD (the dash-dotted line) is such that, by construction, the weighted average of the adjusted industry-specific PDs perfectly matches the aggregate PD. Recall that the aggregate PD is used as an explanatory variable in the industry-specific models. This implies that the industry dynamics of the PDs are broadly consistent with the aggregate dynamics. This desirable feature ensures that the industry models usually behave well (non-negative and non-explosive paths), thereby reducing the need for additional judgement if industry models were estimated less precisely.[25]

[25]Note that the PD of loans to financial institutions is very low and very stable across the business

Figure 9: Projections of industry-specific PD—hypothetical scenario



cycle. Thus, we do not adjust it upward as part of our effort to reconcile the average sectoral PD and the aggregate PD. As a result, the lines "industry original" and "industry adjusted" on Figure 9

6 Conclusion

We present the new corporate default model used in the Bank's macrofinancial stress-testing tool kit. The model forecasts possible default probability paths of the corporate loan portfolio of Canadian banks under extreme but plausible macrofinancial risk scenarios.

Because we can rely on stronger elasticities in the tail of the distribution, our non-linear model can generate higher default probabilities than observed in-sample. Average elasticities estimated from a linear model, in contrast, would be less consistent with the severity of the scenarios typically simulated for stress testing.

Nevertheless, there are a few limitations to be aware of. First, we reconstruct time series of default probabilities, which would otherwise be unavailable over a sufficiently long time span. However, by doing so, we may introduce some model uncertainty related to the specification of the filtering method we use. Second, while our non-linear model helps address the lack of extreme events in our sample, we still need to decide how far out in the tail the elasticities should be estimated.

are very similar for the PD of the loans to financial institutions.

A Data sources and definitions

Table 10: Data sources, coverage and definitions for total loans and gross impaired loans

Corporate Sector's Industries **Source: OSFI A2 Return, starts in 1994Q2, all DSIBs, all currencies**
Natural Resources (sum of (1) Agriculture, (2) Fishing and trapping, and (3) Logging and forestry)
Mining, quarrying and oil wells
Manufacturing
Construction and real estate
Transportation, communications and other utilities
Wholesale trade
Retail trade
Service
Other corporates (sum of (1) Multiproduct conglomerates, (2) Others, and (3) Lease receivables)
Financial institutions
Retail **Source: OSFI N3 Return, starts in 1997Q1, all DSIBs, all currencies**
Consumer Loans (sum of (1) Personal loans, and (2) Other personal credit)
Residential Mortgages (sum of (1) Insured residential mortgages, and (2) Uninsured residential mortgages)
Non-Residential Mortgages (sum of (1) Insured non-residential mortgages, and (2) Uninsured non-residential mortgages)

Note: Corporate Sector's Industries classified under *Standard Industrial Classification* (SIC).

Table 11: Definitions and sources of macrofinancial factors

Category/Variable	**Definition**	**Source**
Canadian		
Economic conditions		
Real GDP	Gross domestic product at market prices (chained 2012 dollars, s.a.a.r)	Statistics Canada
Unemployment Rate	Unemployment rate (both sexes, 15 years and over, s.a.)	Statistics Canada
Credit aggregates		
Private credit-to-GDP ratio	Ratio of the sum of total credit liabilities of households and private non-financial corporations to GDP	Statistics Canada
Business credit-to-GDP ratio	Ratio of total credit liabilities of private non-financial corporations to GDP	Statistics Canada
Real business credit	Total credit liabilities of private non-financial corporations, deflated by consumer price index	Statistics Canada
Residential mortgage-to-household disposable income ratio		Statistics Canada
Consumer credit-to-household disposable income ratio		Statistics Canada
Trade		
Total trade	Sum of total import and export (s.a.a.r)	Statistics Canada
USD/CAD rate	Spot exchange rates (monthly average)	Wall Street Journal
Asset prices and financial conditions		
Corporate spread	ICE BofA Canada Corporate Non-Financial Index	BoA
TSX index	Standard and Poor's/Toronto Stock Exchange Composite Index	Statistics Canada
Real house prices	Residential sale price, average (s.a.)	CREA
Financial stress index	See Duprey (2020)	Bank of Canada
Monetary policy and interest rates		
Inflation Rate	Annual growth rate of Consumer Price Index (s.a.)	Statistics Canada
6-month T-Bills	Treasury bills, 6 month	Statistics Canada
5-year GoC bonds	Selected Government of Canada benchmark bond yields, 5 year	Statistics Canada
Term premium	5-year GoC bonds - 6-month T-Bills	
CA-US MP rate	Bank of Canada Overnight Rate - US Fed Funds Rate	Statistics Canada

Continued on next page

Table 11: continued from previous page

Category/Variable	Definition	Source
Canadian		
Banking		
Total assets	Total assets of the DSIBs	OSFI
Total assets to equity ratio	Ratio of total assets of DSIBs to total shareholders' equity of DSIBs	OSFI
Equity	Total shareholders' equity of DSIBs	OSFI
Return-on-equity	Ratio of net income applicable to common share of DSIBs to total shareholders' equity	OSFI
Net income	Net income applicable to common share	OSFI
US (Foreign)		
Economic conditions		
US Industrial Production	US Total Industrial Production Index (2017=100, s.a.)	Federal Reserve Board
Commodity prices		
Oil price drop	US refiners' acquisition cost of crude oil (including transportation and other fees, dollars per barrel) Computed as $1 - \frac{\text{current oil price}}{\text{maximal oil price over past two years}}$	Energy Information Administration
Asset prices and financial conditions		
US Financial stress index	See Duprey, Klaus, and Peltonen (2017)	Bank of Canada
Monetary policy and interest rates		
US Term premium	US 5-year Gov. Bond - US 3-month T-Bill	Federal Reserve Board

Note: We either use the growth rate in log difference for variables in dollar or the level directly for ratios or rates, except for oil prices that are transformed as the deviation from the local maximum.

B Extension to consumer loans, residential mortgages and non-residential mortgages

We develop an extension that allows simulations of default probabilities for consumer loans, residential mortgages and non-residential mortgages. These extensions are primarily intended to compute the PD distribution required for MFRAF (see section 3.1 of Fique, 2017b), but they can also be used as stand-alone models.

We use the same methodology to develop these models as we did for the corporate sector's default probabilities. First, the PDs for each of these loan types are filtered from the total loans and non-performing loans available in the OSFI N3 return[26] using the methodology described in section 2. Second, we take into account the same macrofinancial predictors that are discussed in section 3.1. Third, the methodology for selecting the most relevant predictors is the same as in section 3.2. We do not, however, perform a new horse race of modelling techniques and use the same ARX and QR models as those selected for the corporate sector. All the results presented in this appendix are for $\lambda = 3$, but the dynamics are similar for $\lambda = \{1, 2, 4\}$.

B.1 Consumer loans

According to the methodology described in this report, the PDs for consumer loans are best explained by:

1. an AR(1) process
2. the current values of the ratio of consumer credit to household disposable income, the TSX index and the US financial stress index
3. the first lagged values of inflation, the policy rate differential between Canada and the United Sates, US industrial production and the growth of oil prices

Figure 10 presents the estimated parameters of the linear autoregressive model (represented by the horizontal dash-dotted black lines) with those of the quantile regressions

[26] See Loans in Arrears (N3) of the Office of the Superintendent of Financial Institutions.

model (represented by the solid black lines), which differ by quantile from 0.1 to 0.9.

Figure 10: Estimated parameters for the PDs for consumer loans—ARX versus QR models

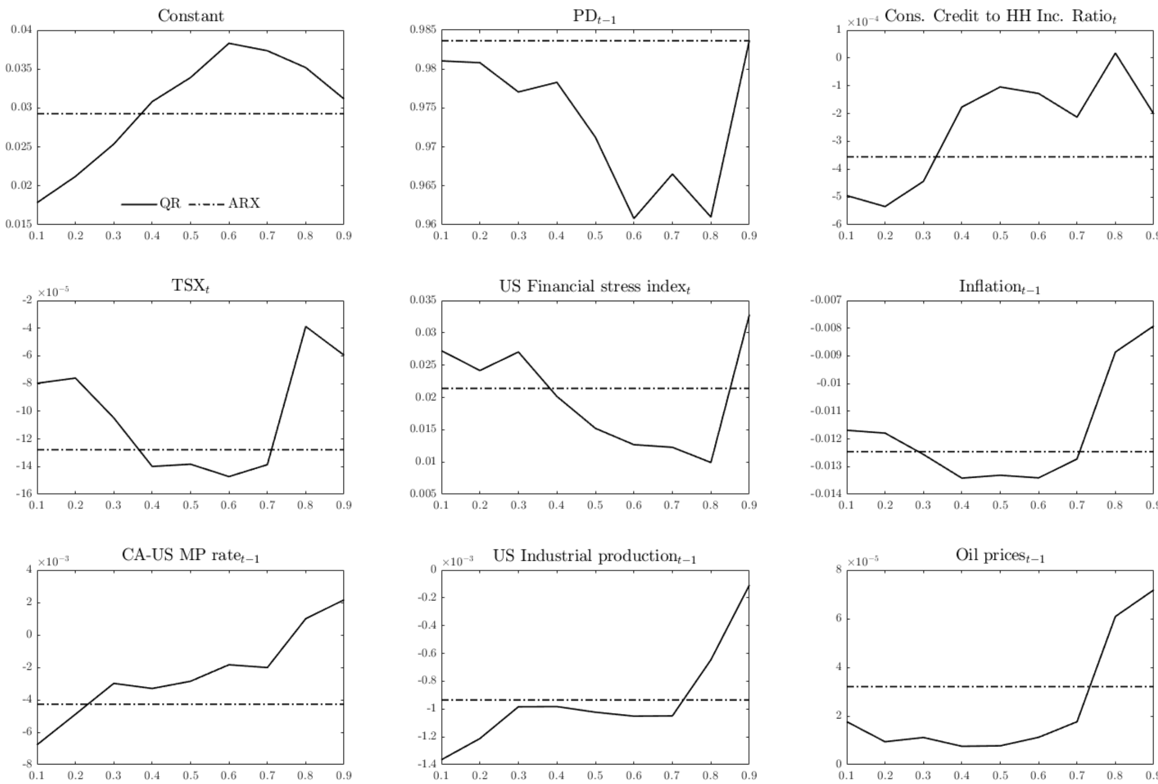
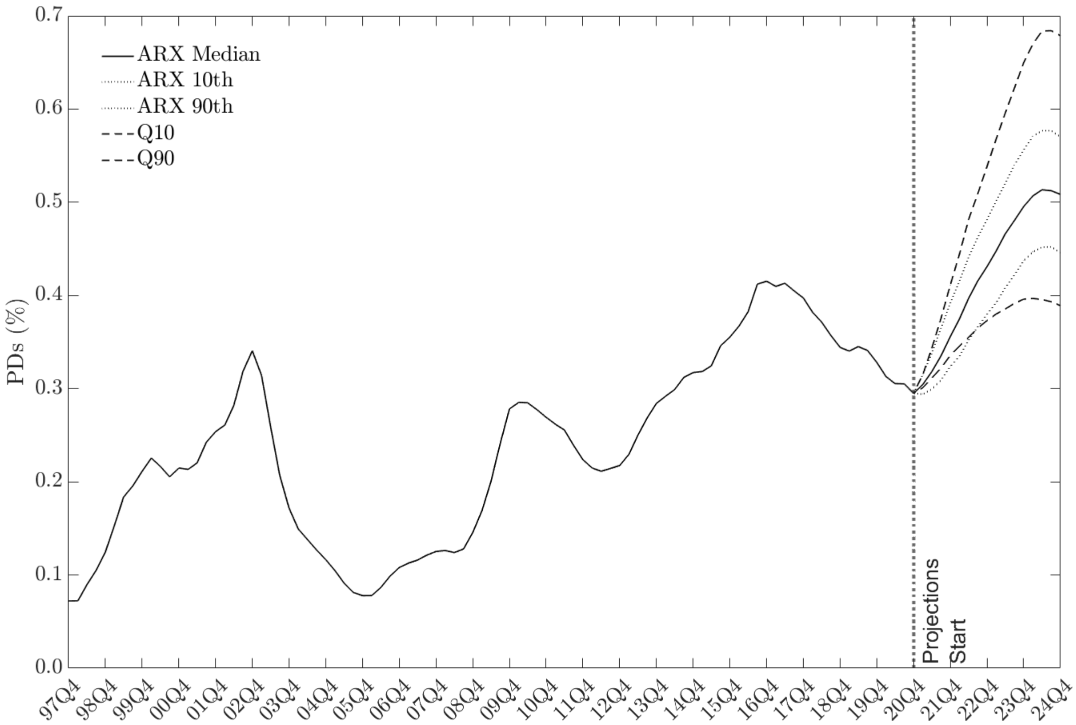


Figure 11 depicts the impact of the risk scenario introduced in section 5 on probabilities of default for the non-residential mortgages on banks' loan portfolios.

Figure 11: Projections of PDs for consumer loans under risk scenario



B.2 Residential mortgages

The PDs for residential mortgages are best explained by:

1. an AR(2) process
2. the current values of the six-month T-bills, the USD/CAD rate, the growth of oil prices, and the term premium
3. the first lagged values of real GDP, the ratio of residential mortgage to household disposable income, the policy rate differential between Canada and the United States, the US financial stress index, and the ratio of business credit to GDP

Figure 12 presents the estimated parameters of the linear autoregressive model (represented by the horizontal dash-dotted black lines) with those of the quantile regressions model (represented by the solid black lines), which differ by quantile from 0.1 to 0.9.

Figure 12: Estimated parameters for the PDs for residential mortgages—
ARX versus QR models

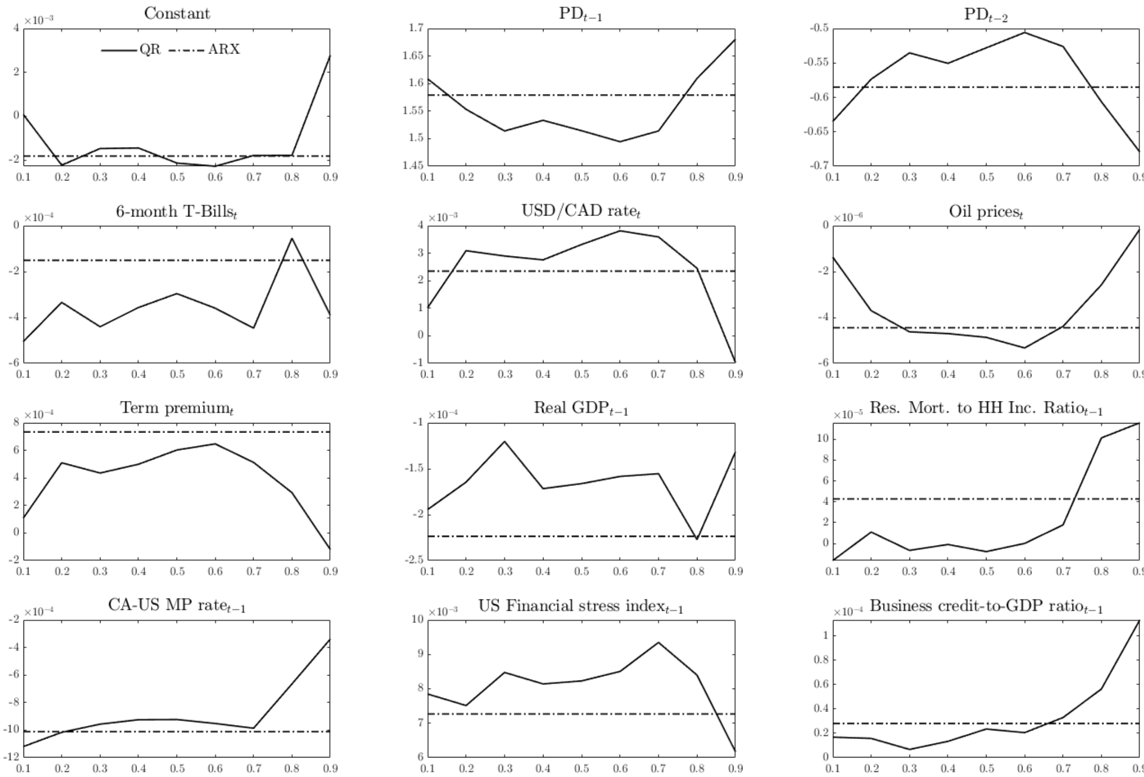
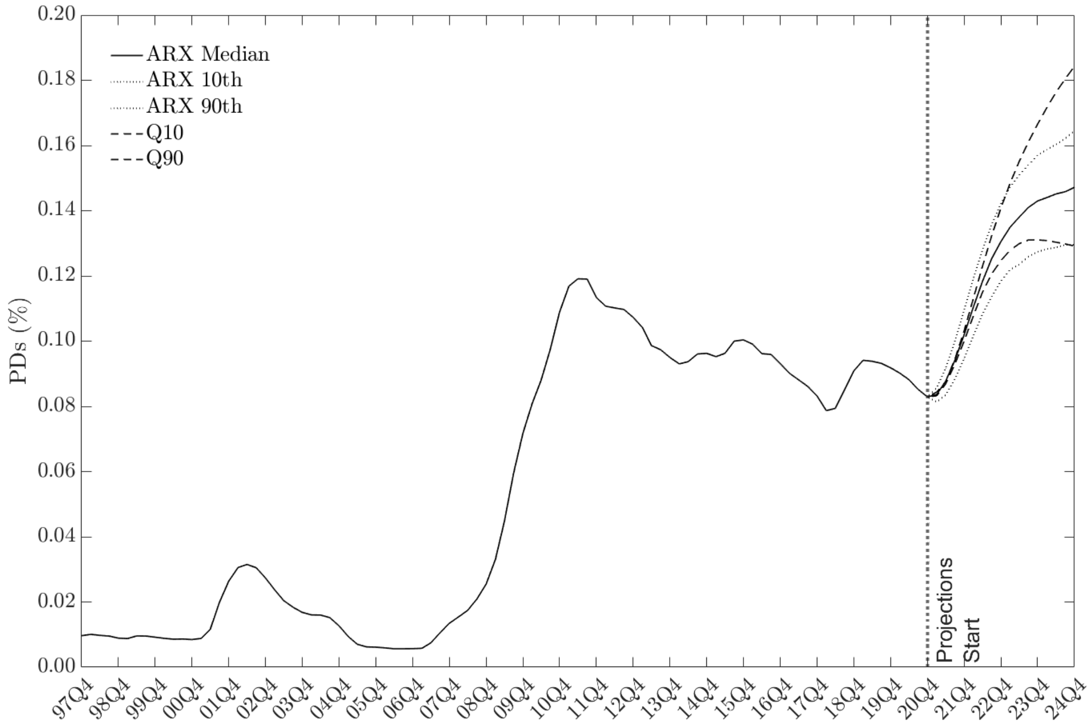


Figure 13 depicts the impact of the risk scenario introduced in section 5 on probabilities of default for the non-residential mortgages on banks' loan portfolios.

Figure 13: Projections of PDs for residential mortgages under the risk scenario



B.3 Non-residential mortgages

The PDs for non-residential mortgages are best explained by:

1. an AR(2) process
2. the current values of inflation, real house prices, the USD/CAD exchange rate and the US financial stress index
3. the first lagged values of the PDs for residential mortgages, the policy rate differential between Canada and the United States
4. the second lagged values of the unemployment rate, and US industrial production

Figure 14 presents the estimated parameters of the linear autoregressive model (represented by the horizontal dash-dotted black lines) with those of the quantile regressions model (represented by the solid black lines), which differ by quantile from 0.1 to 0.9.

Figure 14: Estimated parameters for the PDs of non-residential mortgages—ARX versus QR models

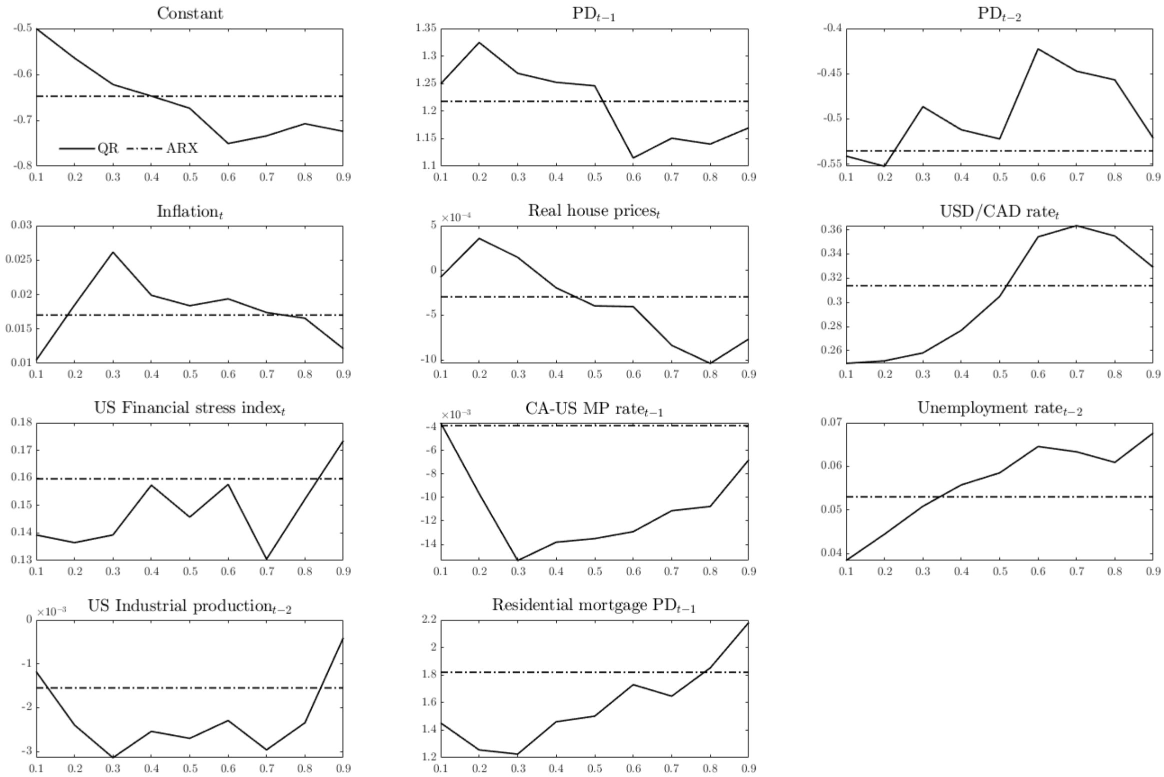
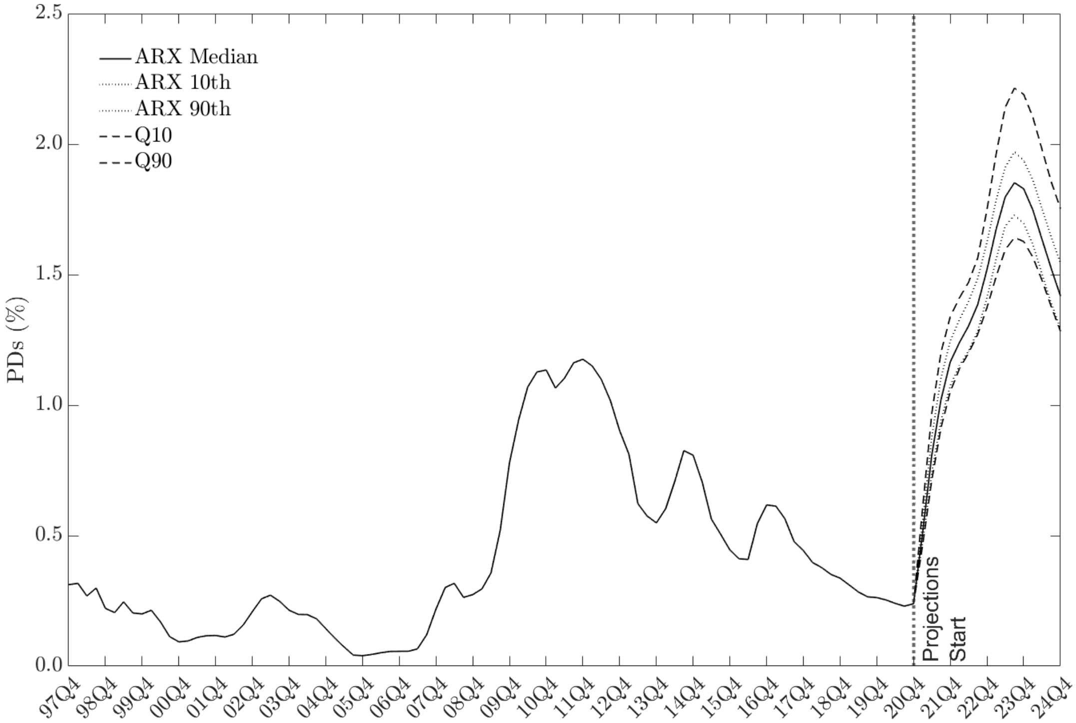


Figure 15 depicts the impact of the risk scenario introduced in section 5 on probabilities of default for the non-residential mortgages on banks' loan portfolios.

Figure 15: Projections of the PDs for non-residential mortgages under the risk scenario



References

Baboucek, I. and M. Jancar. 2005. "Effects of Macroeconomic Shocks to the Quality of the Aggregate Loan Portfolio." Czech National Bank Working Paper No. 1/2005.

Basel Committee on Banking Supervision. 2006. *Sound Credit Risk Assessment for Valuation and Loans.* Basel, Switzerland: Bank for International Settlements.

Beck, R., P. Jakubik and A. Piloiu. 2015. "Key Determinants of Non-performing Loans: New Evidence from a Global Sample." *Open Economies Review* 26 (3): 525–550.

Berger, A. N. and R. DeYoung. 1997. "Problem Loans and Cost Efficiency in Commercial Banks." *Journal of Banking and Finance* 21 (6): 849–870.

Covas, F. B., B. Rump and E. Zakrajšek. 2014. "Stress-Testing US Bank Holding Companies: A Dynamic Panel Quantile Regression Approach." *International Journal of Forecasting* 30 (3): 691–713.

Danielsson, J., M. Valenzuela and I. Zer. 2018. "Learning from History: Volatility and Financial Crises." *Review of Financial Studies* 31 (7): 2774–2805.

De Bock, M. R. and M. A. Demyanets. 2012. "Bank Asset Quality in Emerging Markets: Determinants and Spillovers." International Monetary Fund Working Paper No. 12/71.

Djoudad, R. and É. Bordeleau. 2013. "Méthodologie de construction de séries de taux de défaut pour l'industrie canadienne." Bank of Canada Staff Discussion Paper No. 2013-2.

Duprey, T. 2020. "Canadian Financial Stress and Macroeconomic Condition." *Canadian Public Policy* 46 (S3): S236–S260.

Duprey, T., B. Klaus and T. Peltonen. 2017. "Dating Systemic Financial Stress Episodes in the EU Countries." *Journal of Financial Stability* 32 (C): 30–56.

Espinoza, M. R. A. and A. Prasad. 2010. "Nonperforming Loans in the GCC Banking System and Their Macroeconomic Effects." International Monetary Fund Working Paper No. 10/224.

Figlewski, S., H. Frydman and W. Liang. 2012. "Modeling the Effect of Macroeconomic Factors on Corporate Default and Credit Rating Transitions." *International Review of Economics and Finance* 21 (1): 87–105.

Fique, J. 2017. "The MacroFinancial Risk Assessment Framework (MFRAF), Version 2.0." Bank of Canada Technical Report No. 111.

García-Marco, T. and M. D. Robles-Fernández. 2008. "Risk-Taking Behaviour and Ownership in the Banking Industry: The Spanish Evidence." *Journal of Economics and Business* 60 (4): 332–354.

Godlewski, C. J. 2005. “Bank Capital and Credit Risk Taking in Emerging Market Economies.” *Journal of Banking Regulation* 6 (2): 128–145.

Gross, M. and J. Población. 2019. “Implications of Model Uncertainty for Bank Stress Testing.” *Journal of Financial Services Research* 55 (1): 31–58.

Guth, M. 2022. “Predicting Default Probabilities for Stress Tests: A Comparison of Models.” Working paper.

Hoerl, A. E. and R. W. Kennard. 1970a. “Ridge Regression: Applications to Nonorthogonal Problems.” *Technometrics* 12 (1): 69–82.

Hoerl, A. E. and R. W. Kennard. 1970b. “Ridge Regression: Biased Estimation for Nonorthogonal Problems.” *Technometrics* 12 (1): 55–67.

Hoggarth, G., A. Logan and L. Zicchino. 2005. “Macro Stress Tests of UK Banks.” BIS Papers, No. 22, 392–408.

Hunter, D. R. and K. Lange. 2000. “Quantile Regression via an MM Algorithm.” *Journal of Computational and Graphical Statistics* 9 (1): 60–77.

Jiménez, G. and G. Saurina. 2006. “Credit Cycles, Credit Risk and Prudential Regulation.” *International Journal of Central Banking* 2 (2): 65–98.

Jovovic, J. 2014. “Determinants of Non-performing Loans: Econometric Evidence Based on 25 Countries.” PhD thesis, Master Thesis, City University London.

Kirti, D. 2018. “Lending Standards and Output Growth.” International Monetary Fund Working Paper No. 2018/023.

Koenker, R. and B. J. Park. 1996. “An Interior Point Algorithm for Nonlinear Quantile Regression.” *Journal of Econometrics* 71 (1–2): 265–283.

Louzis, D. P., A. T. Vouldis and V. L. Metaxas. 2012. “Macroeconomic and Bank-Specific Determinants of Non-performing Loans in Greece: A Comparative Study of Mortgage, Business and Consumer Loan Portfolios.” *Journal of Banking and Finance* 36 (4): 1012–1027.

MacDonald, C. and V. Traclet. 2018. “The Framework for Risk Identification and Assessment.” Bank of Canada Technical Report No. 113.

Makri, V., A. Tsagkanos and A. Bellas (2014): “Determinants of Nonperforming Loans: The Case of Eurozone.” *Panoeconomicus* 61 (2): 193–206.

Mileris, R. 2012. “Macroeconomic Determinants of Loan Portfolio Credit Risk in Banks.” *Engineering Economics* 23 (5): 496–504.

Misina, M., D. Tessier and S. Dey. 2006. “Stress Testing the Corporate Loans Portfolio of the Canadian Banking Sector.” Bank of Canada Staff Working Paper No. 2006-47.

Nkusu, M. 2011. "Non-performing Loans and Macrofinancial Vulnerabilities in Advanced Economies." International Monetary Fund Working Paper No. 11/161.

Peterson, B. and T. Roberts. 2016. "Household Risk Assessment Model." Bank of Canada Technical Report No. 106.

Podpiera, J. and L. Weill. 2008. "Bad Luck or Bad Management? Emerging Banking Market Experience." *Journal of Financial Stability* 4 (2): 135–148.

Pouvelle, M. C. 2012. "Bank Credit, Asset Prices and Financial Stability: Evidence from French Banks." International Monetary Fund Working Paper No. 12/103.

Radivojevic, N. and J. Jovovic. 2017. "Examining of Determinants of Non-Performing Loans." *Prague Economic Papers* 26 (3): 300–316.

Rajan, R. and S. C. Dhal. 2003. "Non-performing Loans and Terms of Credit of Public Sector Banks in India: An Empirical Assessment." *Reserve Bank of India Occasional Papers* 24 (3): 81–121.

Rinaldi, L. and A. Sanchis-Arellano. 2006. "Household Debt Sustainability: What Explains Household Non-performing Loans? An Empirical Analysis." European Central Bank Working Paper No. 570.

Saba, I., R. Kouser and M. Azeem. 2012. "Determinants of Non-performing Loans: Case of US Banking Sector." *Romanian Economic Journal* 15 (44): 125–136.

Salas, V. and J. Saurina. 2002. "Credit Risk in Two Institutional Regimes: Spanish Commercial and Savings Banks." *Journal of Financial Services Research* 22: 203–224.

Schularick, M. and A. M. Taylor. 2012. "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008." *American Economic Review* 102 (2): 1029–1061.

Shu, C. 2002. "The Impact of Macroeconomic Environment on the Asset Quality of Hong Kong's Banking Sector." Hong Kong Monetary Authority Research Memorandums (December).

Škarica, B. 2014. "Determinants of Non-performing Loans in Central and Eastern European Countries." *Financial Theory and Practice* 38 (1): 37–59.

Tibshirani, R. 1996. "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–288.

Tuzcuoglu, K. Forthcoming. "The Risk Amplification Macro Model (RAMM)." Bank of Canada Technical Report.

Zou, H. and T. Hastie. 2005. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2): 301–320.