

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 1029

August 2011

Firm Default and Aggregate Fluctuations

Tor Jacobson, Jesper Lindé, and Kasper Roszbach

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at [ww.federalreserve.gov/pubs/ifdp/](http://www.federalreserve.gov/pubs/ifdp/). This paper can be downloaded without charge from Social Science Research Network electronic library at <http://www.ssrn.com>.

Firm Default and Aggregate Fluctuations*

Tor Jacobson

Jesper Lindé

Kasper Roszbach

August 22, 2011

Abstract

This paper studies the relationship between macroeconomic fluctuations and corporate defaults while conditioning on industry affiliation and an extensive set of firm-specific factors. By using a panel data set for virtually all incorporated Swedish businesses over 1990-2009, a period which includes a full-scale banking crisis, we find strong evidence for a substantial and stable impact from aggregate fluctuations on business defaults. A standard logit model with financial ratios augmented with macroeconomic factors can account surprisingly well for the outburst in business defaults during the banking crisis, as well as the subsequent fluctuations in default frequencies. Moreover, the effects of macroeconomic variables differ across industries in an economically intuitive way. Out-of-sample evaluations show that our approach is superior to models that exclude macro information and standard well-fitting time-series models. Our analysis shows that firm-specific factors are useful in ranking firms' relative riskiness, but that macroeconomic factors are necessary to understand fluctuations in the absolute risk level.

Keywords: Default, default-risk model, business cycles, aggregate fluctuations, micro-data, logit, firm-specific variables, macroeconomic variables

JEL: C35, C52, E44, G33.

*Jacobson and Roszbach: Research Division, Sveriges Riksbank, email: firstname.lastname@riksbank.se. Lindé: Federal Reserve Board, email: jesper.linde@frb.gov. We would like to thank Rikard Kindell, who co-authored the working paper version of this paper on a shorter data sample for outstanding contributions to this project. Discussions with and suggestions from Franklin Allen, Ed Altman, Mitch Berlin, Mark Carey, Ines Drumond, Xavier Freixas, Bob Hunt, Wenli Li, Leonard Nakamura, Dragon Tang, Cees Ullersma, and Kostas Tsatsaronis have been very helpful in improving upon earlier drafts. We are also grateful for comments from seminar participants at the Bank of Austria, the Bank of Hungary, the Einaudi Institute for Economics and Finance, the Bank of England, the Bank of Finland, the Federal Reserve Bank of Philadelphia, the Federal Reserve Bank of New York, Uppsala University, EARIE, the C.R.E.D.I.T. 2008 conference, the EEA-ESEM meetings in Budapest (2008) and Barcelona (2009), the 2009 BIS Research task force workshop, the 2008 ASSA meetings, the DNB conference on Financial Stability and Financial Crises, and the BIS. Erica Reisman, Erik von Schedvin and Ingvar Strid provided outstanding research assistance. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank, the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

1 Introduction

The failure of a business is an event of fundamental importance in economic life. Our understanding of the determinants of business defaults, despite long-studying, is far from complete, particularly with respect to the influences of broader economic conditions. Recent economic events, namely a global financial crisis shifting into a recession of exceptional depth, highlight the importance of understanding and predicting this crucial aspect of the economy, not least for timely and appropriate policy measures.

The aim of this paper is to shed light on the dynamics of business defaults. In particular, we seek to understand the interactions between macroeconomic fluctuations and firms' individual likelihoods to default, as well as the relationship between macro variables and the aggregate rate of default. To this end we have compiled a new panel data set with detailed firm-level information on all incorporated Swedish businesses over the period 1990Q1 – 2009Q2. The panel comprises more than 16 million firm-date observations, with an average of more than 200,000 firms per point in time. The length and width of this panel allow for several extensions of previous research; we are able to, among other things, carefully evaluate industry-specific effects of macroeconomic fluctuations. Since our data set includes virtually all incorporated Swedish firms, our findings provide insight into the significance of aggregate fluctuations for both listed and privately held firms; the latter group is typically responsible for over half of GDP in developed economies. This feature is important because Merton-like models of default, being based on equity-price information, are in practice limited to listed firms.

Econometric studies of business defaults started in the 1960s with work by Altman and co-authors (1968, 1971, 1973, 1984). These influential papers focus on explaining bankruptcies of publicly quoted businesses in a cross-sectional context using a small set of firm-specific variables. Later work by Shumway (2001) attempts to account for the dynamic nature of defaults for publicly listed firms. Bharath and Shumway (2008) evaluate the out-of-sample accuracy of the Merton (1974) model and find that the distance-to-default measure is not a sufficient statistic for the probability of default.

Over time the average default frequency and individual default probabilities display co-movement with macroeconomic and financial variables in a way that suggests that aggregate shocks might be an important driver of default. The seminal work of Bernanke, Gertler and Gilchrist (1999) provides a theoretical framework whereby both firm-specific factors and macro shocks affect the default outcome of individual firms. In the BGG framework, firm default is

affected by firm-specific productivity shocks and aggregate shocks (e.g., an aggregate productivity shock). Hence, it follows that an empirical model of default should feature variables that proxy for the underlying firm-specific productivity, as well as variables that proxy for unobserved aggregate shocks. The recent paper by Hackbarth, Miao and Morellec (2006) provides an additional mechanism through which macroeconomic conditions affect default risk. They argue that, when cash flows depend on economic conditions, firms' optimal default thresholds will be affected by aggregate shocks. Hence, aggregate shocks can trigger simultaneous defaults.

These theoretical insights have recently been explored in the empirical literature on default modelling, and there is a small but growing number of papers investigating the importance of macroeconomic fluctuations on business defaults. Recent work by Duffie, Saita and Wang (2007), Pesaran, Schuerman, Treutler, and Weiner (2006), Bonfim (2009), Lando and Nielsen (2010), and Tang and Yan (2010) provides empirical evidence that firm-specific factors alone appear unable to fully explain the variation in corporate default rates and credit spreads. Using aggregate times series and data on publicly listed industrial firms, these authors find that macro-financial covariates have significant explanatory power for credit losses, spreads and corporate default rates using structural and reduced-form approaches.

In this paper, we adopt a standard econometric specification and estimate multi-period logistic regressions on firm-level default data. Shumway (2001) shows that, under some mild restrictions, this model is equivalent to a discrete-time hazard model, and hence not prone to the bias and inconsistency of the static model hitherto used in bankruptcy modelling. In addition to an extensive set of financial statement variables and payment remarks reflecting a firm's financial track record, we include four standard macroeconomic variables. The default risk models are estimated both at an economy-wide level and for 10 industries on a sub-sample covering 1990Q1 – 1999Q4.

Our large panel data set enables us to make several contributions to the above-mentioned literature. First, we show that a simple logit model with constant parameters is able to account for the outburst in default rates during the Swedish banking crisis, as well as the historically low rates occurring in subsequent recovery periods. The included macroeconomic variables are of key importance for explaining the time-varying likelihood of default. Firm-specific variables prove insufficient for explaining variation in the level of default risk over time, but are very useful for ranking firms according to their relative riskiness. Second, having access to a very rich set of firm-specific controls we can credibly reject the possibility that the empirical significance

of macroeconomic variables is merely (or partially) an artifact of a shortage of firm-specific controls. Third, the length of our panel enables us to do extensive out-of-sample performance testing 2000Q1 – 2009Q2. The results suggest that the default risk models with both macroeconomic and firm-specific variables included perform very well out-of-sample. This holds both in the cross-sectional and the time-series dimension. Fourth, the width of our panel permits us to investigate the relationship between aggregate fluctuations and business defaults across industries. By isolating and comparing industry-specific effects from macroeconomic variables we get an additional measure of the robustness of their impact on business defaults. Our results are quite intuitive and suggest that the effects are stronger in sectors like construction and real estate, which on a priori grounds can be deemed to be more cyclical since they involve production of more durable goods. Fifth, the combined width and length of our panel allow us to look into the stability over time of cross-sectionally estimated parameters. We show that models estimated on cross-sectional data are likely to suffer from substantial parameter instability over time, and therefore will be unable to account for variations in the average default frequency over time. Finally, our analysis also suggests that considering only macroeconomic variables while ignoring relevant firm-specific information leads to a substantial loss of out-of-sample prediction accuracy.

According to our analysis, the two key macroeconomic factors affecting business defaults are the nominal interest rate and the output gap. In the current situation, where economic activity and the output gap in many countries have dropped at the fastest pace since the Great Depression, our results suggest that central banks can reduce the likelihood of an outburst in defaults rates and the associated spike in credit losses for banks by aggressively cutting nominal interest rates. Interestingly, this is exactly what many central banks have done during the recent crisis.

The remainder of this paper is structured as follows. The next section presents the micro and macro data sets. The regression results are presented in Section 3 along with an assessment of the in-sample fit. In Section 4, we undertake a thorough out-of-sample investigation of the estimated models along several dimensions. Finally, Section 5 concludes.

2 Data

The firm data set is a panel consisting of 16,928,521 quarterly observations on the population of Swedish aktiebolag, or firms, between January 1, 1990, and June 30, 2009. Aktiebolag are by approximation the Swedish equivalent of US corporations and UK limited liability businesses. Swedish law requires every aktiebolag to have at least SEK 100,000 (approximately US\$ 13,000) of equity to be eligible for registration at Bolagsverket, the Swedish Companies Registration Office (SCRO). Swedish corporations are also required to submit an annual report to the SCRO.

The firm data have been obtained from Upplysningscentralen AB (UC), the leading credit bureau in Sweden, independently operated but jointly owned by the Swedish commercial banks. The UC data come from two general sources. Annual balance sheet and income statement data come from firms' compulsory annual reports submitted to the SCRO. These data cover the period January 1, 1989, through June 30, 2009, and the format follows European Union standards. We convert the annual report data into quarterly observations by linear interpolation, i.e., we assume that the variables remain constant over the quarters in a given reporting period.

The second information source is atypical in the existing default literature and is somewhat unique for Sweden. The credit bureau systematically collects information about events related to firms' payment behavior from all relevant sources, e.g., the Swedish retail banks, the Swedish tax authorities, and the institutions that deal with the legal formalities in firms' bankruptcy processes. The credit bureau thus has a register of more than 60 different payment remarks concerning primarily credit and tax-related events, but also records of various steps in the legal procedures leading up to formal bankruptcy. The information in the register includes a flag for the occurrence of an event in the form of a date and the amount of due payment (if applicable). Some examples of registered events are delays in tax payments, the repossession of delivered goods, the seizure of property, the restructuring of loans, and actual bankruptcy. Payment remarks turn out to be powerful predictors of default and are essentially available in real-time. Admittedly, to allow for comparability with other studies, one might prefer excluding payment remarks, as these are not generally available outside Sweden. However, we prefer to include the payment remarks in our analysis in order to have a comprehensive set of firm-specific control variables. This way, we seek to eliminate the possibility of macroeconomic variables spuriously proxying for omitted firm-specific controls. In Appendix B, however, it is shown that the role of neither the accounting variables nor the macroeconomic factors is much affected by the inclusion of the remark variables that we consider.

The population of existing firms in quarter t is defined as including those firms that have issued a financial statement covering that quarter and are classified as “active,” i.e., the firm has reported total sales and total assets in excess of 1,000 SEK (roughly US\$ 130). However, since there are firms that neglect to fulfill their reporting obligation, a behavior typically associated with distress, we would miss an important segment of firms by only considering those that submit annual reports regularly. For this reason, we will add firms that, according to the data set with payment remarks, are classified as defaulted firms in quarter t . Many firms that default choose not to submit their compulsory annual reports in that year or even for a number of years prior to default. Hence, often the only records of their existence that we have come from the payment remark registers. We adopt the following definition of a default: a firm has default status if any of the following events has occurred: the firm is declared legally bankrupt, has suspended payments, has negotiated a debt composition settlement, is undergoing a re-construction, or is distraint without assets. More details on the construction of the default variable are provided in Appendix A.

For the selection of which financial ratios to use in the default models, we evaluated a large number of frequently used ratios in the literature on bankruptcy risk and on the balance sheet channel.¹ Many papers employ measures of liquidity, profitability and efficiency, and solvency or leverage, while some also make use of a size variable. In this paper the six selected financial ratios are: earnings before interest, depreciation, taxes and amortization (EBITDA) over total assets (TA) (earnings ratio); interest payments (IP) over the sum of interest payments and earnings before interest, depreciation, taxes and amortization (interest coverage ratio); total liabilities (TL) over total assets (leverage ratio); the log of total liabilities over total sales (TS) (debt ratio); liquid assets (LA) in relation to total liabilities (quick ratio); and inventories (I) over total sales (inventory turnover ratio). Details on the selection of financial ratios along with a graphical exposition of the data are provided in Appendix A.2. It is important to note that the non-linear feature of some financial ratios does not imply that these variables are uninformative for default risk when entered linearly in the logit model. The reason for this is that the co-variation between the financial ratios in the cross-section is substantial, which makes each of these variables contribute substantially to predicting default events in the joint linear empirical model. Moreover, the accounting data provides information on whether a firm has paid out

¹ Table A.1 in Appendix A.3 provides an account of the variables considered in Altman (1968), Altman, Frydman, and Kao (1985), Shumway (2001), Pesaran et al (2006), Duffie et al (2007), Bonfim (2009), and Bharath and Shumway (2008).

dividends to shareholders or not, which we enter as a dummy variable (PAYDIV) in the models.

As mentioned previously, some firms classified as active, or defaulted, fail to submit a financial report in every period, leading to a missing-observation problem. For the purpose of using an aggregate default series that closely corresponds to the official default frequency series as computed in the official statistics for Sweden and to ensure unbiased macro coefficients in the econometric model, we decided to retain firms with missing variables in the sample, by replacing missing values by imputation rather than excluding such firms from the sample.² In order to capture the relationship between not submitting a financial statement and subsequent default, we include a dummy variable, denoted TTLFS. In line with actual reporting lags, this dummy equals unity at time t if a firm has not issued a financial statement in the one and a half years prior to the current quarter t , and equals zero otherwise.³ TTLFS attempts to capture the signal that many firms (deliberately) choose not to file a financial report when in financial distress, and thus are more likely to default. In Appendix B.2 we document that our results are robust w.r.t. the approach that we use to deal with missing observations by reporting results where we only include firms for which data on all the financial ratios are available.

For the remark variables, we use simple dummy variables by setting them to unity if certain remarks existed for the firm during the year prior to quarter t , and 0 otherwise. An intuitively reasonable starting point was to find remark events that (i) lead default in time as much as possible and (ii) are highly correlated with default. As it turns out, many payment remark variables are either contemporaneously correlated with default or lack a significant correlation with default behavior. For our final model, we constructed the PAYREMARK variable as a composite dummy of four events: “a bankruptcy petition,” “the issuance of a court order - because of absence during the court hearing - to pay a debt,” “the seizure of property,” and “having a non-performing loan.” The TAXARREARS variable reflects whether the firm is in various tax arrears. It should be emphasized, however, that these two remark variables do not imply a subsequent default incident. The shares of defaulted firms that have received payment remarks, or are in tax arrears, are about 0.15 and 0.41, respectively. Corresponding shares for non-defaulted firms are 0.00 and 0.03. Hence, there are no tautological issues involved in using these variables to predict default events. Table A.2 in Appendix A.3 provides descriptive

² Imputation is implemented in a sequential procedure for a given variable according to: (i) a backward search for the last available observation for that particular firm in the past, (ii) a forward search for future observations for the firm, (iii) if these measures fail, a randomized draw from the data is done conditional on industry and default status.

³ See Appendix A.2 for more details about TTLFS.

statistics for all firm-specific variables that are used in the subsequent analyses.

We make use of four aggregate variables: the output gap (i.e., the deviation of GDP from its trend value), the yearly inflation rate (measured as the fourth difference of the GDP deflator), the repo interest rate (a short-term nominal interest rate, set by the Riksbank), and the real exchange rate. The output gap series is computed by HP-filtering GDP, where the smoothing coefficient λ is set to the standard value of 1,600. The real exchange rate is measured as the trade weighted nominal exchange rate times trade weighted foreign price level (CPI deflators) divided by the domestic CPI deflator. During the sample period, the real exchange rate is characterized by an upward (depreciating) trend and it is therefore detrended with the HP-filter to achieve stationarity. Appendix A.2 provides a figure with the macro data and Appendix B.3 verifies that the results presented below are robust with respect to our detrending procedure of the real exchange rate.

3 The default-risk models: Estimation and in-sample fit

In this section, we examine if default risk at the firm level is affected by aggregate fluctuations over and above the set of firm-specific information that we have at our disposal.

We study the in-sample gains of estimating separate models for each industry and assess the role of aggregate fluctuations for improving the models' fit. The in-sample period is chosen to be 1990Q1 – 1999Q4. For this period, we have a total of 8,106,176 observations of which 105,605 are defaults. The out-of-sample period, 2000Q1 – 2009Q2, is saved to allow for extensive model-evaluation exercises. and comprises a total of 8,822,345 observations of which 55,945 are defaults. Thus, the average default frequency per quarter equals about 1 percent in our dataset for the full sample period. This is somewhat higher than the 0.75 percent per quarter business failure rate reported for the US by, among others, Bernanke, Gertler and Gilchrist (1999), but if we exclude the banking crisis and instead consider the period 1995Q1 – 2009Q2, then the average unconditional default frequency essentially equals the value reported by Bernanke, Gertler, and Gilchrist. Analyses of industry effects will be conducted at the one-digit level to ensure sufficiently many default observations in each industry in both the cross-sectional and the time series dimensions. In addition, we estimate the model for all firms jointly, and will refer to this model as the “economy wide” model.

3.1 The default-risk models

The reduced-form statistical model that we employ for estimating probabilities of default for all Swedish incorporated firms is similar to the multiperiod logit approach used in Shumway (2001) and Campbell, Hilscher, and Szilagyi (2008). Using a reduced-form model both avoids the problem that the Merton (1974) model has, namely that it cannot be implemented for privately held companies without very strong assumptions, and enables us to use a unified approach for all businesses, both privately and publicly held. As discussed in the introduction, there is also a recent theoretical literature - including papers by Bernanke, Gertler and Gilchrist (1999), Hackbarth, Miao and Morellec (2006) and Tang and Yan (2010) - which argues that both firm-specific and aggregate shocks can trigger simultaneous defaults. Thus we propose to estimate the following model:

$$y_{i,t} = x_{i,t}\beta + z_t\gamma + \varepsilon_{i,t},$$

where

$$y_{i,t} = \begin{cases} 1 & \text{if } x_{i,t}\beta + z_t\gamma + \varepsilon_{i,t} \geq 0 \text{ (firm defaults)} \\ 0 & \text{if } x_{i,t}\beta + z_t\gamma + \varepsilon_{i,t} < 0 \text{ (firm stays in business)} \end{cases} ,$$

under the assumption that the vector of firm-specific regressors (i.e., $x_{i,t}$) and the macroeconomic variables we consider (collected in the vector z_t) are stochastically independent with respect to the error term $\varepsilon_{i,t}$. This approach also allows us to control for the competing risk of exiting firms due to other reasons than default (Allison, 1995). We also make the additional assumptions that, conditional on the extensive set of firm-specific and macroeconomic set of covariates we consider, the errors are independent between both firms and time points, i.e., $f(\varepsilon_{i,t}, \varepsilon_{j,t}) = f(\varepsilon_{i,t})f(\varepsilon_{j,t})$ for $i \neq j$ and $f(\varepsilon_{i,t}, \varepsilon_{i,t+p}) = f(\varepsilon_{i,t})f(\varepsilon_{i,t+p})$ for $p \neq 0$. These assumptions are rejected by Das, Duffie, Kapadia and Saita (2007) on US data. Lando and Nielsen (2010), however, revisit the relation between contagion through covariates and conditional dependence addressed in Das et al. and find that the assumption of conditional independence can no longer be rejected when the set of firm-specific and macroeconomic controls is slightly altered and expanded. Hence, conditional on an appropriate set of covariates, Lando and Nielson find no evidence that the default of a firm causes default intensities of other firms to increase, providing support for our assumptions for the error term. Moreover, for a similar model estimated on a subset of the data used in this paper, Carling, Roszbach and Rönnegård (2004) find that the estimated parameters are robust when the correlation between residuals is taken into account. These results, along with the excellent out-of-sample performance of the model in the time and cross-section dimension

documented in Section 4, provide support of our assumptions for the error term.

Our theoretical basis for selecting the set of macroeconomic variables is that they should span both aggregate “demand” and “supply” type of shocks that hit the economy. The output gap is intended as an indicator of demand conditions, i.e., increased demand in the economy is expected to reduce default risk. We also include the nominal interest rate in z_t because credit conditions facing firms, in particular firms in distress, are likely to be tightly linked to the interest rate. As one can plausibly argue that the real interest rate, rather than the nominal one, is what should affect the default frequency, we also include the inflation rate in z_t . Moreover, apart from capturing the effects of supply shocks, higher inflation obviously implies higher nominal income for firms, which furthermore should tend to reduce default risk. However, it is also conceivable that higher inflation rates are associated with less certainty about correct relative prices and thus may lead to increased default risks. For these reasons, the sign of the inflation coefficient is unclear, and will depend on the relative strength of the underlying sources of macroeconomic fluctuations. Furthermore, given that the export-to-GDP ratio in Sweden was around 0.40 during the sample period, the real exchange rate is potentially an important variable, since a depreciation renders improved competitiveness to Swedish firms in the export sector.⁴

3.2 Estimation results

To document how aggregate variables contribute to the default risk models, we present estimation results for two specifications: one with and one without macroeconomic variables. Moreover, results are presented for ten industry-specific models and an economy-wide model where firms in all industries are jointly modelled.

Table 1 contains estimation results for a model with firm-specific determinants of default risk only (i.e., the six financial ratios augmented with the dummy variables PAYDIV, TTLFS, PAYREMARK, and TAXARREARS), while Table 2 shows results with the macroeconomic variables added. The regressors have not been re-scaled to have the same mean, and therefore one cannot directly judge the importance of a particular variable from the size of its coefficient. However, by suitably transforming each variable, its marginal impact is calculated in Appendix B.4, verifying that such calculations yield similar rankings of importance as standard t -statistics. Hence, the importance of each variable is below approximated with the size of its t -statistic.

⁴ In addition to these four variables, we have also experimented with a few other variables such as real housing prices, taken as deviations from linear/HP-trend; and a measure of the spread between the interest rate charged to non-financial firms and the policy rate set by the central bank. For our sample period, these variables are largely redundant given the set of included variables in the benchmark specification.

Since the firms’ annual financial reports are typically submitted with a significant time lag, it cannot in general be assumed that accounting data for year τ are available during, or even at the end of, year τ and enable forecasted default risks for year $\tau + 1$. To account for this, all accounting data is lagged by four quarters in the estimations. For most firms, which report balance-sheet and income-statement data over calendar years, this means that data for year τ are assumed to have been available in the first quarter of year $\tau + 1$. It should be emphasized that our decision to lag the accounting data four quarters in the estimation in order to make the model “operational” in real time has minor implications for the estimated coefficients. When re-estimating the model using contemporaneous data instead, the estimation results were found to be very similar to the ones reported in Tables 1 and 2.

The results in Table 1 show that the firm-specific information we consider is indeed important for explaining default behavior in both the industry-specific models and in the economy-wide model. In particular, the indicator variable TTLFS (which takes a value of 1 if a firm has not filed an annual report on time, and 0 otherwise) and the variables for remarks on firms’ payment records are very powerful predictors of default. Among the financial ratios we find the earnings ratio EBITDA/TA, leverage ratio TL/TA and the debt ratio TL/TS to be quite useful. However, the roles played by financial ratios in the various industry models differ substantially; while accounting data are less important in the financial services (bank, finance and insurance) sector, it is more important in the manufacturing industry. The coefficients for the payment remarks and the indicator variable TTLFS are quite similar across industries. So to the extent that these variables are the more important ones for explaining firm default behavior, there is no clear gain at the firm-specific level from conditioning on industry.⁵

Turning to the results in Table 2, we find that the coefficients for the firm-specific variables in Table 1 do not change much when the model is augmented with the macroeconomic variables. Moreover, and despite the robustness of the firm-specific coefficients, we find that all coefficients for the macroeconomic variables are significant in the economy-wide model and have the expected signs. The possible exception is inflation, but for reasons discussed in Section 3.1, it is hard to have a strong view a priori on the sign of the inflation coefficient. The notion that it is important to condition on macroeconomic variables in default risk modeling is further supported by the industry-specific model results. Table 2 shows that the impact of the macroeconomic

⁵ Notice that by defining a default event at the quarterly frequency, and by transforming yearly statements to quarterly ones, we could potentially underestimate the effects of the accounting variables. As a robustness check we therefore estimated the default-risk models on annual data and found that the coefficients for the accounting variables are quite similar to those reported in Table 1, see results in Appendix B.3.

factors is estimated to be more important in industries that are arguably more cyclical. For instance, the output gap is more important in the construction and the real estate sectors in comparison with other industries, and as expected the nominal interest rate is found to be very important for the financial services and the real estate sectors. The remaining macroeconomic variables, inflation and the real exchange rate, appear less important overall. However, it is reassuring to find that a depreciating real exchange rate (i.e., increasing values) is associated with a significantly lower default risk in the manufacturing sector, which is the most export-oriented industry. The coefficient for the real exchange rate is also large for the financial services industry, possibly reflecting that Swedish credit conditions, which were very tight in the resolvment of the banking crises, subsequently eased when the krona was allowed to depreciate in November 1992. Regarding inflation, we can reject the view that it is only the real interest rate that matters for default risk at the firm level, with the possible exception of the financial services industry.⁶

Finally, we would like to emphasize that the gain in using firm-specific data for default-risk modeling is substantial. OLS estimation (TSLS with lagged variables as instruments yielded very similar results) for a model of the average quarterly default rate on average financial ratios and the four macroeconomic variables yields:

$$\begin{aligned}
df_t = & \begin{matrix} -9.45 \\ (10.67) \end{matrix} \begin{matrix} -0.27 \\ (0.18) \end{matrix} \left(\frac{\text{EBITDA}}{\text{TA}} \right)_t + \begin{matrix} 0.17 \\ (0.12) \end{matrix} \left(\frac{\text{TL}}{\text{TA}} \right)_t - \begin{matrix} 0.003 \\ (0.02) \end{matrix} \left(\frac{\text{LA}}{\text{TL}} \right)_t \dots \\
& \begin{matrix} -0.40 \\ (0.17) \end{matrix} \left(\frac{\text{I}}{\text{TS}} \right)_t \begin{matrix} -0.02 \\ (0.04) \end{matrix} \left(\frac{\text{TL}}{\text{TS}} \right)_t + \begin{matrix} 0.16 \\ (0.16) \end{matrix} \left(\frac{\text{IP}}{\text{IP+EBITDA}} \right)_t \dots \\
& \begin{matrix} -0.12 \\ (0.06) \end{matrix} y_{d,t} - \begin{matrix} 0.007 \\ (0.04) \end{matrix} \pi_{d,t} - \begin{matrix} 0.005 \\ (0.04) \end{matrix} R_{d,t} - \begin{matrix} 0.02 \\ (0.01) \end{matrix} q_t + \hat{u}_{df,t}, \tag{1}
\end{aligned}$$

$$R^2 = 0.88, \text{ DW} = 1.98, \text{ Sample: } 1990Q1 - 1999Q4, (T = 40).$$

If we compare the point estimates for the financial ratios in (1) with the economy-wide model in Table 2, we see that they differ substantially and the ratios I/TS and TL/TS appear with counterintuitive signs. The coefficients for the macroeconomic variables are more robust, with the exception of the nominal interest rate which has a counterintuitive sign. Since the average financial ratios are quite smooth over time, it is not surprising that we obtain spurious results when the firm-specific information is aggregated. Moreover, we notice that some explanatory

⁶ As a robustness check, we examined a model allowing for non-linear relationships between default and the financial ratios and found that the macroeconomic variables are still highly significant and quantitatively important. We used the cumulated distributions depicted in Figure A.1 in Appendix A to categorize the variables (3 categories for each variable). For instance, we classified EBITDA/TA into the decile-based categories 0 – 10, 10 – 90, 90 – 100, whereas TL/TA was classified into the categories 0 – 75, 75 – 90, 90 – 100. This categorization resulted in an increase in pseudo R^2 from 0.34 to 0.42 in the economy-wide Table 2 model, but the macroeconomic variables still enter highly significantly and with coefficients very close to those reported in Table 2.

power is lost by aggregating data; the model in (1) yields an R^2 of 0.88, which can be directly compared with the aggregated fit ($R^2 = 0.95$) of the corresponding model in Table 2. The reduction in fit is primarily driven by the inability to take advantage of the dummy variables for payment remarks, dividends and failure to submit a financial statement in regressions at the aggregate level.

3.3 Assessing the models' in-sample fit

The last rows in Tables 1 and 2 report on the number of observations, the mean log-likelihood and the pseudo- R^2 . The latter measures the ability of the estimated models to explain default at the firm level and is computed using the method of McFadden (1974). Another important and interesting feature of the models is their aggregate performance over time, i.e., how well the models account for the average default frequency. Hence, we report what we label as “industry” or “aggregate” R^2 s. These are calculated by aggregating all the fitted firm default probabilities in a particular industry model for each quarter 1990Q1 – 1999Q4 and then using the resulting 40 time-series observations to compute the implied aggregate R^2 . To assess the gain in estimating separate industry-specific models, we also report the pseudo- and industry- R^2 values conditional on the economy-wide model coefficients instead of the industry model coefficients.

By comparing Tables 1 and 2, we see that the pseudo- R^2 is not much affected by conditioning on macroeconomic factors in any of the industries, merely 1-2 percentage points. Tang and Yan (2010) find a somewhat larger role for macro factors: about 6 percentage points. However, the industry- R^2 is doubled and sometimes even more than doubled by the introduction of macroeconomic variables. Thus, the firm-specific variables account for the cross-section of the default distribution, while the macroeconomic variables in the model play the role of shifting the mean of the default distribution in each period. This also implies that the model with firm-specific information cannot capture the upturns and downturns in the average default rate over time. This is visualized in Figure 1, where we plot the average default rate over time against the fitted values from the economy-wide models with (Table 2) and without (Table 1) macroeconomic variables. The results to the right-hand side of the vertical line pertain to out-of-sample results and will be discussed in Section 4.1. According to Figure 1, the model with both microeconomic and macroeconomic variables included indeed appears able to replicate the extreme default rates during the deep recession/banking crisis in the beginning of the 1990s, as well as the downturn to moderate default rates towards the end of the in-sample period. This finding is very inter-

esting, because it suggests that the extreme default rates recorded during the banking crisis in the early 1990s were not exceptional events that are uninformative in a model context. Rather, they appear to be consequences of unusually bad economic outcomes. An additional feature of interest in Tables 1 and 2 is that the fall in pseudo- R^2 values associated with conditioning on the economy-wide model coefficients is distinct but limited, whereas the corresponding reduction in industry R^2 is often quite substantial. In three cases (the agricultural, the bank, finance & insurance, and the non-classified sectors) the industry- R^2 are negative conditional on the economy-wide model coefficients. At first sight this may seem strange, given that the industry-specific coefficients in Table 2 are not very different from the economy-wide model coefficients. However, these seemingly inconsistent results are driven by the unreported intercept, which is larger in the economy-wide model compared with the sector models. Therefore, it induces a systematic over-prediction of default risk in these sectors.

A conceivable objection to our claim of the importance of conditioning on macroeconomic factors in default risk models is that the significance of these variables simply reflects a changing impact of the firm-specific variables over time. Accordingly, if one were to continuously re-estimate the coefficients of the models in Table 1 using the most recent quarterly information only, then the fit of the models in terms of aggregate R^2 would increase dramatically and make the macroeconomic variables redundant. Figure 2 displays the estimated coefficients for the financial ratios in the economy-wide model for such a set of separate cross-sectional models.⁷ These results are computed for the economy-wide model only, because there are not enough defaults available in each quarter to estimate industry-specific models. As can be seen from Figure 2, the coefficients for most ratios are highly unstable and some even switch sign over time. Accordingly, any out-of-sample forecasts beyond the very short horizon generated by any of these 40 models would be deficient. To be convincing, a model of firm-level defaults with time-varying coefficients for firm-specific variables would require an understanding of how the time-variation in its coefficients comes about. The irregular and economically implausible patterns in Figure 2, make such a model seem highly improbable.

To further understand the role of macroeconomic variables for default risk, let us approach the issue from an opposite angle and study the importance of firm-specific variables in the

⁷ We have also conducted these cross-sectional regressions when imposing the restriction that the constant equals the estimated intercept in the economy-wide model in Table 1 and subtracting the panel mean from each regressor. So instead of running the regressions underlying Figure 2, i.e. $y_{i,t} = \alpha_t + x'_{i,t}\beta_t + e_{i,t}$ for each quarter, we estimated $y_{i,t} = \bar{\alpha} + (x'_{i,t} - \bar{x}')\beta_t + e_{i,t}$. This alternative estimation procedure yielded very similar results to those reported in Figure 2.

models. One way to demonstrate the information loss due to omitting the microdata is to regress the average default frequency exclusively on the macroeconomic variables. When doing so we obtain the following result:

$$df_t = \underset{(0.11)}{0.40} - \underset{(0.03)}{0.20}y_{d,t} + \underset{(0.02)}{0.007}\pi_{d,t} + \underset{(0.02)}{0.10}R_{d,t} - \underset{(0.009)}{0.03}q_t + \hat{u}_{df,t},$$

$$R^2 = 0.81, \text{ DW} = 1.43, \text{ Sample: } 1990Q1 - 1999Q4, (T = 40). \quad (2)$$

When comparing this regression with the economy-wide model in Table 2, we see that exclusion of the financial statement variables is associated with a loss of close to 15 percentage points of explanatory power. Moreover, omitting the firm-specific information introduces misspecification problems in (2) as indicated by the DW-statistic, in contrast to the results in (1), which has a DW-statistic around 2 and hence displays no signs of autocorrelation. A simple F -test reveals that the loss of fit in (2) relative to (1) is significant at the 5-percent level (using asymptotic critical values). The autocorrelation problem in (2) turns out to induce further problems with out-of-sample stability for the model in (2), as documented in Section 4 (see Table 3). Our interpretation is that omitting firm-specific variables when modeling default risk attributes too much of the variation in default risk to the macroeconomic factors in-sample. The model therefore doesn't perform as well out-of-sample.

4 Out-of-sample performance of the estimated model

In this section we investigate the robustness of the results in the previous section by examining the out-of-sample performance of the models of Tables 1 and 2 for the period 2000Q1 – 2009Q2. We evaluate the models along two dimensions. First, we study the models' performance at the industry and aggregate level, i.e., we assess their ability to predict future average default rates. The predictions we consider are static one-step-ahead forecasts because we do not have a complete model for all the regressors. Although there are no major fluctuations in the aggregate default rate during the out-of-sample period (see Figure 1), this period is nevertheless very informative about the out-of-sample performance of our models since there is still substantial variation in the macroeconomic variables, as displayed in Figure A.2. Second, we look into the models' properties in predicting future default events at the firm level. To this end, there is a substantial amount of information (55,945 default observations) to assess the stability of the models.

4.1 Evaluating the models at the aggregate and the industry level

In Figure 1, the results to the right-hand side of the vertical bar show the one-step-ahead, out-of-sample performance at the aggregate level for the economy-wide model. Overall, the out-of-sample fit is remarkably good, although there is a tendency for the model to under-predict during the years 2006 and 2007. Interestingly, the model is able to pick up the emerging spike in the default frequency in the recent recession through the strong fall in the output gap from 4 to -4 percent.

In Table 3, we report the root mean squared one-step-ahead prediction errors (RMSEs), for the estimated models of Tables 1 and 2. As a reference, we also show results for three reference time series models: a random walk, a 4-quarter moving-average model, and the model estimated on only macroeconomic data (eq. 2, denoted “Industry OLS macroregression”). The results in Table 3 pertain to default risk models that have been re-estimated using macroeconomic variables that are lagged one quarter. This ensures that all models in Table 3 have been estimated on the same information set, thereby allowing for a fair comparison between the logit and the time-series models. In the “Industry OLS macroregression” models an additional dummy for the third quarter is included.

In order to assess to which extent the forecasting errors are quantitatively different from a statistical point of view, we perform the Diebold-Mariano test on the forecast errors underlying the computed RMSE differentials in the lower panel of Table 3. In the table the RMSE-ratios that are bolded (in italics) indicate that the forecasting performance is significantly better (worse) relative to the corresponding models in Table 2.⁸ Finally, it is imperative to notice that the RMSEs are shown in percent, i.e., the actual and fitted default frequencies have been multiplied with a factor of 100 before the prediction errors are calculated.

From inspection of Table 3, it is evident from the first row in the lower panel that the effect on forecasting performance from conditioning on both macro and firm-specific information is considerable. The largest gain is found for the economy-wide model where forecast precision increases by a factor of 3.6 when we include macroeconomic variables. Disregarding the not classified residual industry for which the model with macro variables is associated with a significant loss of out-of-sample accuracy, the corresponding factors for the industry-specific models

⁸ Our application of the test suggested by Diebold and Mariano (1995) examines the null hypothesis: $E[L(\varepsilon_{t+1|t}^i)] = E[L(\varepsilon_{t+1|t}^j)]$, where L is the squared loss-function of the one-step-ahead forecast errors $\varepsilon_{t+1|t}$ for models i and j . Diebold and Mariano show that a test-statistic based on the loss differential d_t , and suitably normalized by the asymptotic variance of d_t , is asymptotically standard normal.

range between 1.2 and 3.6 and constitute a significant improvement in 8 out of 9 cases. Moreover, the second row in the lower panel shows that the industry-specific models often generate significantly lower RMSEs compared with the industry models conditional on coefficients from the economy-wide model in Table 2, except for the manufacturing, retail and hotel & restaurant sectors where they are slightly, but not significantly, higher.

By and large, the above findings constitute evidence that the industry-specific models are not over-parameterized with respect to the macroeconomic variables. Therefore it will typically be worthwhile to work with an industry-specific model if the focus is on understanding default behavior in a particular industry. However, if interest lies in modeling aggregate default behavior only, then the economy-wide default model appears to suffice. This tentative conclusion can be drawn from the two right-most columns of the second row in Table 3. In the second to last column, the forecasts computed with the industry-specific models have been weighted (according to industry-size) to a forecast for the aggregate default frequency. This results in a slightly lower RMSE (0.0716) in comparison with the RMSE (0.0802) for the economy-wide model. Although this difference in RMSE is moderate in comparison with the other models in absolute terms, it is still significant in favor of the industry models according to the D-M test.

Comparing the industry models in Table 3 with the time series models, we see that while the random walk model is doing significantly better in 3 out of 10 sectors, and the 4-quarter moving-average specification is better 6 out of 10 times at the industry level, they are both significantly inferior at the aggregated industry level. This implies that they are also inferior in terms of RMSE fit to the economy-wide model specification in Table 2 (which conditions on aggregate fluctuations). The models that are based on OLS regressions for average industry default frequencies on the macroeconomic variables only are often associated with a significant increase in RMSE (7 out of 10 industries) in comparison with the Table 2 models.

To sum up, we have found strong evidence that the favorable fit in-sample of the estimated industry (and aggregate) models, conditional on macroeconomic variables, is preserved out-of-sample at the industry and aggregate level. This is reassuring for the hypothesis that aggregate variables matter because the in-sample and the out-of sample periods taken together cover several upturns and downturns in the Swedish economy. Finally, we have also documented that there are relatively small gains in terms of forecasting accuracy to be made by using industry-specific models rather than simply an economy-wide model, as long as an appropriate set of macroeconomic variables is included.

4.2 Evaluating the models at the firm and the industry level

We now turn to an evaluation of the ability of the models to both rank firms according to their relative riskiness and to determine firms' absolute risk level for the out-of-sample period. In addition, we report the industry-specific pseudo- R^2 conditional on the industry-specific model coefficients of Table 2, as well as the pseudo- R^2 calculated conditional on the economy-wide model coefficients. The results are displayed in Table 4.

First, starting with the pseudo- R^2 for the models with industry-specific coefficients and comparing the in-sample and out-of-sample results reported in Tables 2 and 4, respectively, we see that the explanatory power out-of-sample is in fact either higher than in-sample or unchanged in eight out of ten industries.

Next we turn to the pseudo- R^2 for the predictions based on the economy-wide model coefficients. The lower panel of Table 4 shows that the average explanatory power has increased substantially, from 0.34 in-sample to 0.38 out-of-sample. We also see, relative to Table 2, that the explanatory power has increased in all industries except for agriculture, real-estate and the construction sectors. Moreover, by comparing the pseudo- R^2 values generated when using industry-specific coefficients with those obtained using economy-wide model coefficients (i.e. the values in upper and lower panels of Table 4), we find that they are larger/smaller in four industries and similar otherwise. This implies that pseudo- R^2 at the aggregate level is about the same (0.38) for the economy-wide model compared with an aggregation of pseudo- R^2 over the industry-specific models (denoted Industry Aggregate in Table 4). These results provide support for two important conclusions. First, the industry models are not over-parameterized. Second, the reduced-form coefficients appear to be stable over time and the regressions thus reflect relationships that hold out-of-sample.

Moving on to measures of relative risk, we follow Shumway (2001) and evaluate the models' ability to rank firms according to their riskiness in terms of ex post default frequencies. At first glance, we see from Table 4 that the estimated models classify roughly 75 – 80 percent of the defaulting firms in the riskiest decile. These numbers are about the same as those reported in-sample by Shumway for a data set that was substantially smaller and included only listed firms. Our models cover the entire population of Swedish incorporated businesses, of which only a very small subset is publicly listed on the stock exchange (slightly less than 500 out of 260,000). We therefore conclude that our models are quite successful in ranking firms according to their level of default risk, and support our conclusion that the role of macroeconomic variables in our

models of default risk is not driven by the omission of key microeconomic variables.

Table 4 also reveals that the quality of the risk rankings does not depend on whether we condition on industry-specific coefficients or coefficients from the economy-wide model. This contrasts with our findings in the previous subsection, where we found that conditioning on industry-specific parameters improved the models' empirical performance at the industry level. The explanation for these seemingly inconsistent results lies in the fact that the most important difference between the economy-wide and industry-specific models is due to the varying impact of the aggregate factors. Those factors have little impact on the firms' relative risk ranking and hence their inclusion or omission has little impact on the models' ability to risk-rank firms.

Finally, we assess the out-of-sample properties of the models at the microeconomic level in an absolute sense, as opposed to the relative appraisal in Table 4. We do so by sorting all estimated default probabilities according to size and calculate the average probability of default in each percentile. We then compare the average probabilities of default with the actual default experience of the firms for each of these percentiles. In Figure 3, we plot the result where we have used both the industry-specific and the economy-wide model coefficients in Tables 1 and 2 to compute the estimated default probabilities for each firm. On the x-axis, we have the estimated default frequency in a given percentile, and on the y-axis, we have the actual default frequency in each percentile. In the figure, each dot represents a percentile. In order to make the results easier to interpret, a logarithmic scale is used for both the estimated and the actual series. If the estimated models could perfectly predict the absolute riskiness of the firms within each percentile, all dots would line up along the 45-degree line drawn in the figures, corresponding to a slope-coefficient of unity and an intercept equal to zero. As can be seen in Figure 3, this is not the case for either model, but the dots are generally very close to the line, suggesting that the absolute riskiness ranking is very accurate. In particular, the results show that both the industry-specific models and the economy-wide model with macroeconomic variables included pass our test in the cross-sectional dimension, since they are not systematically below or above the 45-degree line, whereas the models without macroeconomic variables tend to overestimate default risk. These findings provide further support for the main theme of this paper; macroeconomic variables are key for getting the absolute risk level right, but are not important for ranking firms according to their relative riskiness in a given period.

5 Conclusions

In this paper, we studied the interaction between macroeconomic fluctuations and default risk at the firm level using reduced-form methods; we present five main findings. First, we provide insight into the significance of aggregate fluctuations for defaults among both listed and privately held firms. This is important, since privately held businesses typically account for over half of GDP in developed economies. Second, a nearly exhaustive set of firm-specific background variables permits us to investigate the importance of and interaction between firm-specific variables and macroeconomic information - an area that so far has received little attention. Third, we document that a standard logit approach to model default at the firm level, using both firm-specific and macroeconomic variables, can explain the extreme default frequencies during the Swedish banking crisis of the early 1990s as well as the considerably lower default frequencies in the late 1990s. Fourth, the estimated models are shown to be very robust and successful out-of-sample, suggesting that aggregate fluctuations play an important role in understanding the absolute level of firm default risk. Fifth and finally, the width of our panel permits us to investigate the relationship between aggregate fluctuations and firm defaults across industries. This shows that macroeconomic variables have a robust impact on business defaults.

We want to stress that we do not interpret our results to imply that aggregate fluctuations are the most important source of default risk at the firm level. Rather, we argue that the results suggest that macroeconomic factors shift the mean of the default risk distribution over time and thereby are the most important determinants of the average level of default.

In view of these results, we conclude by providing some suggestions as to why aggregate fluctuations have an important impact on firm default behavior, over and above the effects of firm-specific variables, which themselves move in response to macroeconomic fluctuations. We have in this paper relied on the work of, among others, Bernanke, Gertler and Gilchrist (1999) and Hackbarth, Miao and Morellec (2006) to motivate why firm default should be affected by aggregate shocks. In addition, one could imagine several additional channels whereby aggregate variables might contain predictive information for firm-default risk over and above firm-specific information. One such explanation is related to the costliness of monitoring. If monitoring borrowers is costly for banks, then banks may use aggregate information to assess the probability of getting repayment on loans granted. That is, banks may form their credit-granting policies on the basis of macroeconomic forecasts and decide to not extend new lines of credit to firms with a given set of performance indicators in one particular phase of the business cycle, but

readily do so in another phase. The tightening and loosening of banks' credit standards over various phases of the business cycle captures such behavior.⁹ Yet another argument follows a similar line of reasoning. If entrepreneurs have imperfect information about their own future business prospects, they may resort to using aggregate conditions as a basis for their decision to either invest more effort in a firm or declare bankruptcy. A final possibility is that firms may be inclined to adjust their yearly accounts, e.g., to smooth profit over time in order to please banks' monitoring efforts, and thereby reduce the predictive power of firm-level information. We believe that further work on the theory of how macroeconomic variables affect firm defaults and assessing the empirical relevance of the arguments above are important issues for future research.

References

- Allison, Paul D., (1995), *Survival Analysis Using SAS. A Practical Guide*, SAS Publishing.
- Altman, Edward I., (1968), Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, 23(4), pp. 589-611.
- Altman, Edward I., (1971), Railroad bankruptcy propensity, *Journal of Finance*, 26(2), pp. 333-345.
- Altman, Edward I., (1973), Predicting railroad bankruptcies in America, *Bell Journal of Economics*, 4(1), pp. 184-211.
- Altman, Edward I., (1984), The success of business failure prediction models: An international survey, *Journal of Banking and Finance*, 8(2), pp. 171-198.
- Altman, Edward I., and Anthony Saunders, (1997), Credit risk measurement: developments over the last twenty years, *Journal of Banking and Finance*, 21(11-12), pp. 1721-42.
- Altman, Edward, Halina Frydman, and Duen-Li Kao, (1985), Introducing recursive partitioning for financial classification: the case of financial distress, *Journal of Finance*, XL(1), pp. 269-291.

⁹ The empirical results in Lown and Morgan (2006) and Jiménez, Ongena, Peydró and Saurina (2011) support this reasoning.

- Bernanke, Ben S., Mark Gertler, and Simon Gilchrist (1999), "The Financial Accelerator in a Quantitative Business Cycle Framework." Chapter 21 in Handbook of Macroeconomics, Volume 1, edited by J.B. Taylor and M. Woodford. Amsterdam, Elsevier.
- Bharath, Sreedhar T., and Tyler Shumway, (2008), Forecasting default with the Merton distance to default model, *Review of Financial Studies* 21(3), pp. 1339-1369.
- Bonfim, Diana, (2009), Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics , *Journal of Banking and Finance*, 33, pp. 281-299.
- Campbell, John, Jens Hilscher and Jan Szilagyi, (2008), In Search of Distress Risk, *Journal of Finance*, 63(6), pp. 2899-2939.
- Carling, Kenneth, Kasper Roszbach and Lars Rönnegard (2004), Is Firm Interdependence within Industries Important for Portfolio Credit Risk?, *Sveriges Riksbank Working Paper Series*, No 168.
- Das, S., D. Duffie, N. Kapadia, and L. Saita (2007). Common failings: How corporate defaults are correlated, *Journal of Finance*, 62, 93-117
- Diebold, Francis, and Roberto Mariano, (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, pp. 253-265.
- Duffie, Darrell, Leandro Saita and Ke Wang, (2007), Multi-period corporate default prediction, *Journal of Financial Economics*, 83(3), pp. 635-665.
- Hackbarth, Dirk, Jianjun Miao and Erwan Morellec, (2006), Capital structure, credit risk and macroeconomic conditions, *Journal of Financial Economics*, 82(3), pp. 519-550.
- Jiménez, Gabriel, Steven Ongena, José Luis Peydró and Jesús Saurina (2011), Credit Supply and Monetary Policy: Identifying the Bank Balance-Sheet Channel with Loan Applications, *American Economic Review*, forthcoming.
- Lando, David, and Mads Stenbo Nielsen, (2010), Correlation in Corporate Defaults: Contagion or Conditional Dependence?, *Journal of Financial Intermediation*, 19, pp. 355-372.
- Lown, Cara and Donald Morgan, (2006), The Credit Cycle and the Business Cycle: New Findings Using the Loan Officer Opinion Survey, *Journal of Money, Credit and Banking*, 38(6), pp. 1575-1597.

- McFadden, Daniel, (1974), The measurement of urban travel demand, *Journal of Public Economics*, 3(4), pp. 303-328.
- Merton, Robert C., (1974), On the pricing of corporate debt: the risk structure of interest rates, *Journal of Finance*, 29, pp. 449-470.
- Pesaran, M. Hashem, Til Schuermann, Björn-Jakob Treutler, and Scott M. Weiner, (2006), Macroeconomic dynamics and credit risk: a global perspective, *Journal of Money, Credit and Banking*, 38(5), pp. 1211-1261.
- Shumway, Tyler, (2001), Forecasting bankruptcy more accurately: a simple hazard model, *Journal of Business*, 74(1), pp. 101-124.
- Tang, Dragon Yongjun, and Hong Yan, (2010), Market conditions, default risk and credit spreads, *Journal of Banking and Finance*, 24, pp. 743-753.

Table 1: Regression results 1990Q1-1999Q4 for the default risk model estimated with only firm-specific variables

	Agriculture	Manu- facturing	Construction	Retail	Hotel & Restaurant	Transport	Bank. Finance & Insurance	Real Estate	Consulting & Rental	Not Classified	Economy Wide
Firm-specific variables ^a											
EBITDA/TA	-1.004 (0.094)	-1.123 (0.038)	-1.195 (0.044)	-0.825 (0.021)	-0.599 (0.034)	-1.031 (0.055)	-0.329 (0.068)	-0.755 (0.059)	-0.806 (0.026)	-0.758 (0.026)	-0.837 (0.011)
TL/TA	0.693 (0.061)	0.835 (0.028)	0.452 (0.032)	0.471 (0.013)	0.240 (0.023)	0.840 (0.046)	0.138 (0.036)	0.472 (0.028)	0.292 (0.021)	0.244 (0.020)	0.395 (0.007)
LA/TL	-0.254 (0.067)	-0.390 (0.029)	-0.307 (0.028)	-0.272 (0.014)	-0.203 (0.033)	-0.154 (0.032)	-0.085 (0.028)	-0.089 (0.020)	-0.182 (0.012)	-0.117 (0.010)	-0.207 (0.006)
I/TS	0.047 (0.034)	0.177 (0.030)	-0.064 (0.027)	0.131 (0.014)	0.241 (0.237)	0.472 (0.153)	-0.062 (0.032)	0.034 (0.009)	0.239 (0.031)	0.207 (0.020)	0.057 (0.005)
TL/TS	0.118 (0.021)	0.092 (0.005)	0.187 (0.007)	0.086 (0.004)	0.058 (0.011)	0.015 (0.010)	0.081 (0.015)	0.086 (0.007)	0.118 (0.006)	0.215 (0.006)	0.108 (0.002)
IP/(IP+EBITDA)	0.102 (0.032)	0.098 (0.011)	0.060 (0.013)	0.047 (0.006)	0.014 (0.017)	0.154 (0.024)	0.044 (0.043)	0.162 (0.019)	0.042 (0.012)	0.066 (0.013)	0.066 (0.004)
PAYREMARK	1.336 (0.111)	1.431 (0.041)	1.685 (0.043)	1.557 (0.027)	1.568 (0.058)	1.687 (0.063)	2.456 (0.148)	1.787 (0.060)	1.826 (0.038)	2.583 (0.051)	1.726 (0.014)
TAXARREARS	2.905 (0.072)	2.308 (0.026)	2.492 (0.027)	2.483 (0.017)	2.410 (0.039)	2.804 (0.041)	2.913 (0.102)	2.543 (0.039)	2.845 (0.025)	2.479 (0.029)	2.565 (0.009)
PAYDIV	-3.471 (0.709)	-2.912 (0.168)	-3.021 (0.190)	-3.268 (0.133)	-2.458 (0.335)	-3.406 (0.410)	-3.690 (1.004)	-3.187 (0.355)	-2.878 (0.159)	-3.578 (0.260)	-3.173 (0.071)
TTLFS	4.013 (0.066)	3.513 (0.024)	3.931 (0.027)	3.543 (0.015)	3.326 (0.038)	3.913 (0.040)	3.618 (0.085)	3.665 (0.032)	3.773 (0.022)	3.906 (0.025)	3.694 (0.008)
Mean log-likelihood	-0.024	-0.042	-0.045	-0.052	-0.071	-0.036	-0.025	-0.051	-0.033	-0.076	-0.046
Pseudo R ²	0.38	0.28	0.35	0.31	0.30	0.38	0.38	0.33	0.37	0.42	0.33
Pseudo R ² agg.par. ^b	0.36	0.27	0.34	0.31	0.30	0.37	0.30	0.33	0.37	0.41	0.33
Industry R ²	0.32	0.48	0.46	0.49	0.47	0.44	0.20	0.59	0.45	0.68	0.42
Industry R ² agg.par. ^b	-4.88	0.46	0.39	0.46	0.40	0.29	-4.22	0.59	0.14	-0.82	0.51
Number of obs	260,941	1,242,335	914,337	2,260,316	263,648	502,158	120,995	444,609	1,611,700	485,137	8,106,176

Notes: Standard errors in parentheses. The variables are not scaled, so the importance of a variable cannot be interpreted directly from the size of the parameter estimate. ^a See Section 2 for definitions of these variables. ^b Pseudo R² | agg.par is the Pseudo R² value calculated for each industry using the estimated coefficients in the economy-wide model (i.e., the coefficients in the last column in the table above). The pseudo R² values are calculated according to McFadden (1974). In addition to the coefficients reported above, three more variables were included (but not reported). First, an industry-specific intercept. Second, since the bankruptcy rate is systematically lower in the third quarter (most likely due to Swedish courts' summer holiday period in July-August), a seasonal dummy is included to capture this phenomenon. Third, because no data on the payment records of firms (i.e., the dummy variables PAYREMARK and TAXARREARS) exist prior to 1992Q3 for legal storage reasons, the models also include one additional variable common to all *i* firms that is constructed to be an estimate of the average value of the sum of the payment record variables PAYREMARK and TAXARREARS for the quarters 1990Q1-1992Q2. This variable was constructed by estimating a logit model for the event of either of the dummy variables PAYREMARK and TAXARREARS taking on the value 0 or 1 for the period 1992Q3-1999Q2, using all the other variables in the model in Table 1 as regressors (except PAYREMARK and TAXARREARS, of course). The imputed average value for this variable for the period 1990Q1-1992Q2 (after 1992Q2, it is set to nil) was then constructed as the average estimated probability for each firm and period.

Table 2: Regression results 1990Q1-1999Q4 for the default risk model estimated with both firm-specific and aggregate variables

	Agriculture	Manu- facturing	Construction	Retail	Hotel & Restaurant	Transport	Bank, Finance & Insurance	Real Estate	Consulting & Rental	Not Classified	Economy Wide
Firm-specific variables ^a											
EBITDA/TA	-1.005 (0.093)	-1.092 (0.038)	-1.119 (0.043)	-0.794 (0.021)	-0.584 (0.034)	-1.016 (0.055)	-0.312 (0.068)	-0.675 (0.058)	-0.811 (0.027)	-0.753 (0.026)	-0.822 (0.011)
TL/TA	0.685 (0.062)	0.817 (0.028)	0.432 (0.033)	0.476 (0.013)	0.244 (0.023)	0.811 (0.046)	0.124 (0.037)	0.488 (0.029)	0.264 (0.021)	0.230 (0.020)	0.383 (0.007)
LA/TL	-0.270 (0.067)	-0.391 (0.029)	-0.303 (0.028)	-0.278 (0.014)	-0.208 (0.033)	-0.158 (0.033)	-0.083 (0.028)	-0.102 (0.021)	-0.182 (0.012)	-0.116 (0.010)	-0.204 (0.006)
I/TS	0.022 (0.035)	0.174 (0.030)	-0.075 (0.027)	0.119 (0.014)	0.245 (0.236)	0.468 (0.157)	-0.075 (0.032)	0.024 (0.009)	0.189 (0.032)	0.184 (0.020)	0.044 (0.005)
TL/TS	0.111 (0.021)	0.084 (0.005)	0.182 (0.007)	0.073 (0.004)	0.040 (0.011)	0.004 (0.011)	0.087 (0.015)	0.073 (0.007)	0.110 (0.006)	0.210 (0.006)	0.101 (0.002)
IP/(IP+EBITDA)	0.095 (0.031)	0.086 (0.011)	0.052 (0.012)	0.039 (0.006)	0.006 (0.017)	0.130 (0.023)	0.044 (0.042)	0.139 (0.019)	0.036 (0.011)	0.060 (0.013)	0.056 (0.004)
PAYREMARK	1.449 (0.113)	1.558 (0.042)	1.841 (0.043)	1.676 (0.027)	1.668 (0.058)	1.828 (0.064)	2.539 (0.150)	1.940 (0.061)	1.950 (0.038)	2.688 (0.051)	1.850 (0.015)
TAXARREARS	2.983 (0.073)	2.423 (0.027)	2.647 (0.028)	2.602 (0.017)	2.494 (0.040)	2.904 (0.042)	3.014 (0.104)	2.639 (0.041)	2.989 (0.026)	2.587 (0.030)	2.684 (0.009)
PAYDIV	-3.352 (0.709)	-2.728 (0.168)	-2.806 (0.190)	-3.095 (0.134)	-2.299 (0.335)	-3.232 (0.410)	-3.509 (1.003)	-2.968 (0.355)	-2.714 (0.159)	-3.459 (0.261)	-3.007 (0.071)
TTLFS	3.954 (0.067)	3.421 (0.025)	3.822 (0.027)	3.446 (0.016)	3.228 (0.039)	3.848 (0.040)	3.598 (0.086)	3.500 (0.033)	3.700 (0.022)	3.849 (0.025)	3.608 (0.008)
Aggregate variables ^b											
Output gap	-0.130 (0.021)	-0.131 (0.007)	-0.200 (0.008)	-0.126 (0.005)	-0.141 (0.011)	-0.138 (0.012)	-0.137 (0.031)	-0.171 (0.011)	-0.133 (0.007)	-0.070 (0.008)	-0.133 (0.003)
Nominal interest rate	0.066 (0.013)	0.062 (0.004)	0.083 (0.005)	0.070 (0.003)	0.048 (0.007)	0.060 (0.007)	0.096 (0.018)	0.115 (0.007)	0.079 (0.004)	0.060 (0.005)	0.073 (0.002)
GDP inflation	-0.043 (0.018)	0.016 (0.006)	-0.020 (0.007)	0.020 (0.004)	0.037 (0.010)	0.020 (0.011)	-0.084 (0.027)	0.000 (0.009)	0.013 (0.006)	0.027 (0.007)	0.012 (0.002)
Real exchange rate	-0.021 (0.006)	-0.034 (0.002)	-0.030 (0.002)	-0.023 (0.001)	-0.021 (0.003)	-0.031 (0.004)	-0.034 (0.009)	-0.027 (0.003)	-0.031 (0.002)	-0.021 (0.002)	-0.026 (0.001)
Mean log-likelihood	-0.024	-0.042	-0.044	-0.051	-0.070	-0.035	-0.025	-0.050	-0.032	-0.076	-0.046
Pseudo R ² .	0.38	0.28	0.36	0.32	0.30	0.38	0.38	0.35	0.38	0.43	0.34
Pseudo R ² agg.coeffs. ^c	0.38	0.28	0.36	0.32	0.31	0.38	0.32	0.34	0.38	0.41	0.34
Industry R ²	0.75	0.92	0.91	0.95	0.89	0.86	0.71	0.87	0.93	0.93	0.95
Industry R ² agg.coeffs. ^c	-4.95	0.90	0.85	0.92	0.72	0.66	-4.86	0.83	0.47	-0.48	0.95
Number of obs	260,941	1,242,335	914,337	2,260,316	263,648	502,158	120,995	444,609	1,617,700	485,137	8,106,176

Notes: Standard errors in parentheses. The variables are not scaled, so the importance of a variable cannot be interpreted directly from the size of the parameter estimate. ^a See Section 2 for definitions of these variables. ^b See Section 2.2 for definitions. ^c Pseudo R² | agg.coeffs. is the Pseudo R² value calculated for each industry using the estimated coefficients in the economy-wide model (i.e., the coefficients in the last column in the table above). The pseudo R² values are calculated according to McFadden (1974). In addition to the coefficients reported above, three more variables were included (but not reported). First, an industry-specific intercept. Second, since the bankruptcy rate is systematically lower in the third quarter (most likely due to Swedish courts' summer holiday period in July-August), a seasonal dummy is included to capture this phenomenon. Third, because no data on the payment records of firms (i.e., the dummy variables PAYREMARK and TAXARREARS) exist prior to 1992Q3 for legal storage reasons, the models also include one additional variable common to all i firms that is constructed to be an estimate of the average value of the sum of the payment record variables PAYREMARK and TAXARREARS for the quarters 1990Q1-1992Q2. This variable was constructed by estimating a logit model for the event of either of the dummy variables PAYREMARK and TAXARREARS taking on the value 0 or 1 for the period 1992Q3-1999Q2, using all the other variables in the model in Table 1 as regressors (except PAYREMARK and TAXARREARS, of course). The imputed average value for this variable for the period 1990Q1-1992Q2 (after 1992Q2, it is set to nil) was then constructed as the average estimated probability for each firm and period.

Table 3: Out-of-Sample Root Mean Squared Error (RMSE) for various models

Model	RMSE (in percent) ^b											
	Agriculture	Manu- facturing	Construction	Retail	Hotel & Restaurant	Transport	Bank, Finance & Insurance	Real- Estate	Consulting & Rental	Not Classified	Industry aggregate	Economy Wide
Absolute RMSE for Model j ^a												
Firm-specific and macro	0.0821	0.1013	0.1145	0.1148	0.3110	0.0735	0.1200	0.1322	0.0664	0.7843	0.0716	0.0802
Only firm-specific variables	0.1848	0.2986	0.3725	0.4103	0.7034	0.2569	0.1456	0.4516	0.2342	0.4411	0.2856	0.2902
Economy-wide coefficients	0.2516	0.0874	0.1173	0.1058	0.2740	0.0830	0.1863	0.2732	0.1419	1.5866	0.0802	0.0802
Time series random walk ^d	0.0877	0.1297	0.1361	0.1418	0.2102	0.1489	0.1467	0.0734	0.1013	0.6084	0.1365	0.1371
Industry OLS macroregression ^d	0.1514	0.1124	0.1879	0.2107	0.4808	0.1582	0.1263	0.3708	0.1084	0.7933	0.1320	0.1308
4 quarter moving average ^d	0.0703	0.1241	0.1035	0.1068	0.1598	0.1143	0.1289	0.0674	0.0858	0.4921	0.1031	0.1033
RMSE model j / RMSE												
Table 2 model ^c												
Only firm-specific variables	2.2509	2.9477	3.2533	3.5740	2.2617	3.4952	1.2133	3.4160	3.5271	<i>0.5624</i>	3.9888	3.6185
Economy-wide coefficients	3.0646	<i>0.8628</i>	1.0245	0.9216	<i>0.8810</i>	1.1293	1.5525	2.0666	2.1370	2.0230	1.1201	1.0000
Time series random walk ^d	1.0682	1.2804	1.1886	1.2352	<i>0.6759</i>	2.0259	1.2225	<i>0.5552</i>	1.5256	<i>0.7757</i>	1.9064	1.7095
Industry OLS macroregression ^d	1.8441	1.1096	1.6410	1.8354	1.5460	2.1524	1.0525	2.8048	1.6325	1.0115	1.8436	1.6309
4-quarter moving average ^d	0.8563	1.2251	0.9039	0.9303	<i>0.5138</i>	1.5551	1.0742	<i>0.5098</i>	1.2922	<i>0.6274</i>	1.4399	1.2880

Notes: The RMSEs have been computed as one-step-ahead forecasts for the period 2000Q1-2009Q2. The RMSE ratios have been computed relative to the first row in the upper panel, i.e., the industry-specific models. All models were estimated for the period 1990Q1-1999Q4. Industry aggregate RMSEs have been computed by summing the default frequency probabilities implied by each industry model quarterly. ^a Note that the macro variables in these forecasting models are lagged one quarter, so that all models are based on the same information set. ^b Notice that the RMSE numbers are expressed in percent, i.e., fitted and actual default numbers are multiplied by 100 before the RMSE numbers are computed. ^c Values in bold (italics) style indicate that the RMSE-ratios are significantly better (worse) according to the Diebold-Mariano test at the 5 percent level (two-sided test). The Diebold-Mariano test is computed under the assumption of no serial correlation in the RMSE differentials. ^d These models are estimated on industry or economy-wide data, and not at the firm level.

Table 4: Out-of-sample Pseudo R² and decile tests at the industry level

Industry-specific model coefficients											
	Agriculture	Manufacturing	Construction	Retail	Hotel & Restaurant	Transport	Bank. Finance & Insurance	Real-Estate	Consulting & Rental	Not Classified	Industry aggregate
Pseudo R²	0.35	0.29	0.41	0.33	0.35	0.39	0.44	0.26	0.40	0.46	0.38
Decile											
1	0.79	0.74	0.85	0.76	0.80	0.82	0.81	0.77	0.81	0.79	0.81
2	0.09	0.11	0.06	0.09	0.07	0.08	0.04	0.06	0.07	0.08	0.06
3	0.05	0.06	0.03	0.05	0.03	0.04	0.04	0.04	0.04	0.04	0.04
4	0.02	0.04	0.02	0.04	0.02	0.02	0.02	0.03	0.03	0.03	0.03
5	0.02	0.02	0.01	0.02	0.02	0.01	0.05	0.03	0.02	0.02	0.02
6 - 10	0.04	0.04	0.03	0.04	0.06	0.03	0.04	0.08	0.03	0.04	0.04
Sum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Economy-wide model coefficients											
	Agriculture	Manufacturing	Construction	Retail	Hotel & Restaurant	Transport	Bank. Finance & Insurance	Real-Estate	Consulting & Rental	Not Classified	Aggregate
Pseudo R²	0.28	0.30	0.41	0.35	0.34	0.39	0.37	0.28	0.37	0.47	0.38
Decile											
1	0.79	0.72	0.85	0.75	0.79	0.81	0.81	0.76	0.81	0.80	0.78
2	0.08	0.10	0.06	0.08	0.08	0.07	0.04	0.05	0.07	0.08	0.07
3	0.05	0.07	0.03	0.06	0.03	0.03	0.04	0.04	0.04	0.04	0.05
4	0.02	0.04	0.02	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03
5	0.02	0.02	0.01	0.02	0.02	0.02	0.03	0.03	0.02	0.02	0.02
6 - 10	0.04	0.04	0.03	0.04	0.06	0.04	0.04	0.09	0.03	0.04	0.04
Sum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Notes: The out-of-sample period is 2000Q1-2009Q2, and the total number of firms in the panel for this period is 8,822,345. The decile test outcomes in the table are obtained by sorting the estimated default probabilities, in descending order, and by computing observed default frequencies in the different deciles of the sorted data. The coefficients used for calculating the default probabilities are the ones presented in Table 3. Industry aggregate numbers are obtained by generating the estimated default probabilities using industry-specific model coefficients, then aggregating observations over the various industries to form a single data set, to which, finally, the procedure outlined above is applied in order to compute default frequencies for various deciles.

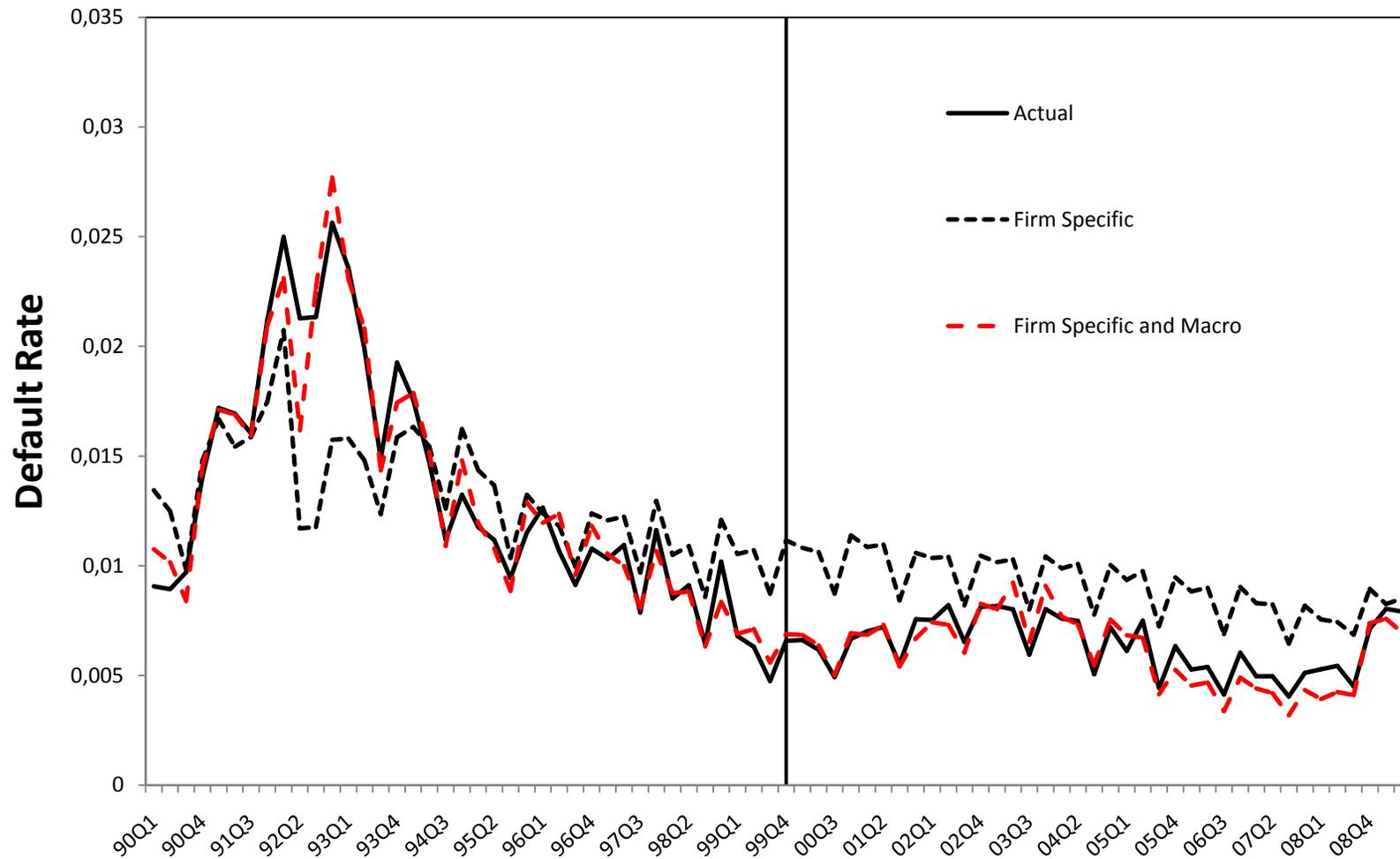


Figure 1: Actual (solid) and projected (dashed, dotted) aggregate default frequency rates 1990Q1-2009Q2. The projected rates are constructed using the estimated economy-wide models in Table 1 (dotted) and Table 2 (dashed). The models are estimated on data until 1999Q4. The projections shown to the right of the vertical line are out-of-sample.

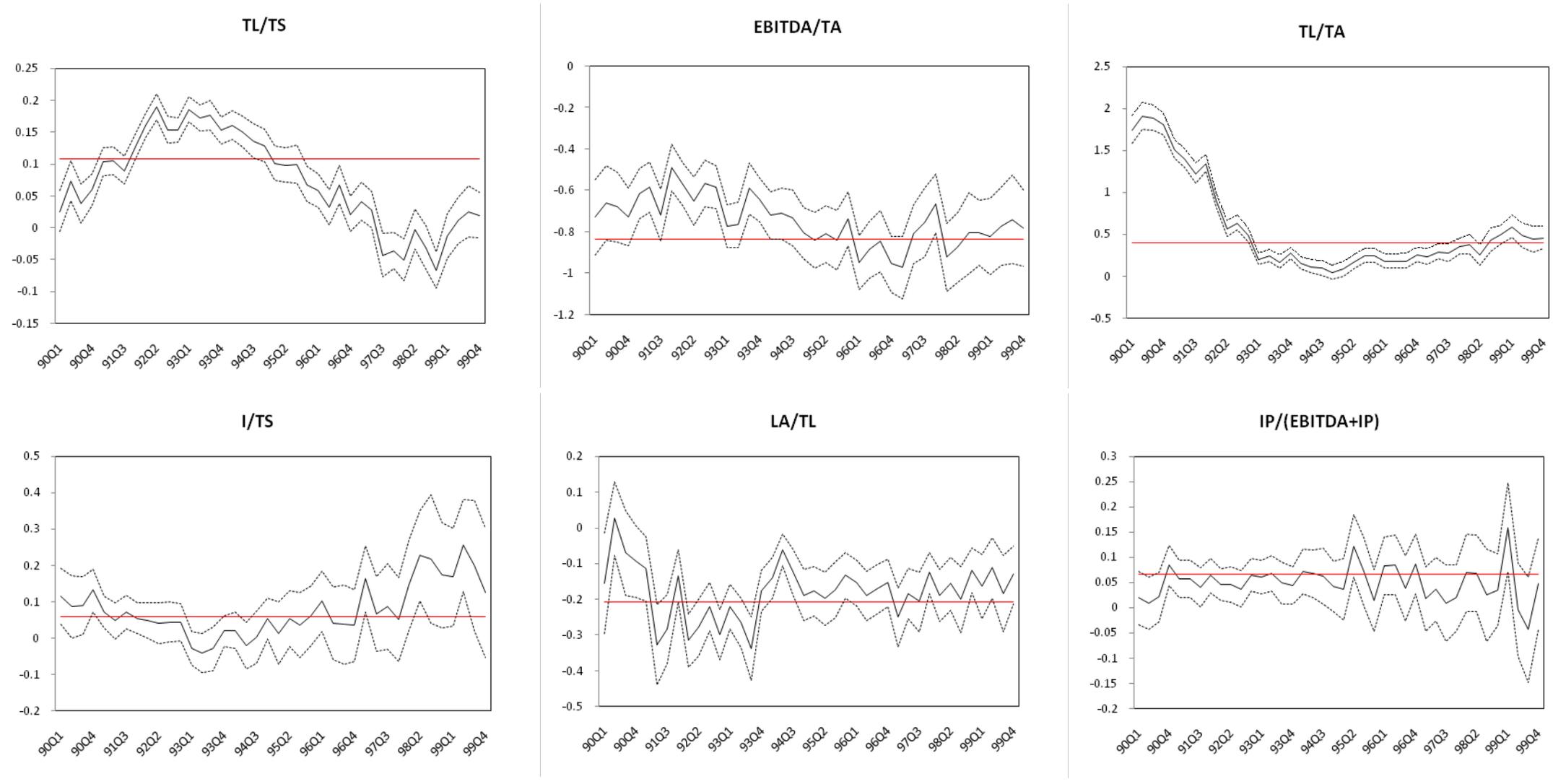


Figure 2: Time-varying coefficients for the financial ratios in the economy-wide model

Notes: Coefficients have been calculated by estimating a separate economy-wide model for each quarter up to 1999Q4. The macroeconomic variables have been excluded from the model. Intercepts are allowed to vary by quarter. Solid lines are coefficient estimates, dashed lines are 95-percent confidence bands based on the standard errors from each quarterly regression. The solid (red) horizontal lines correspond to the estimated coefficients in the economy-wide model in Table 1.

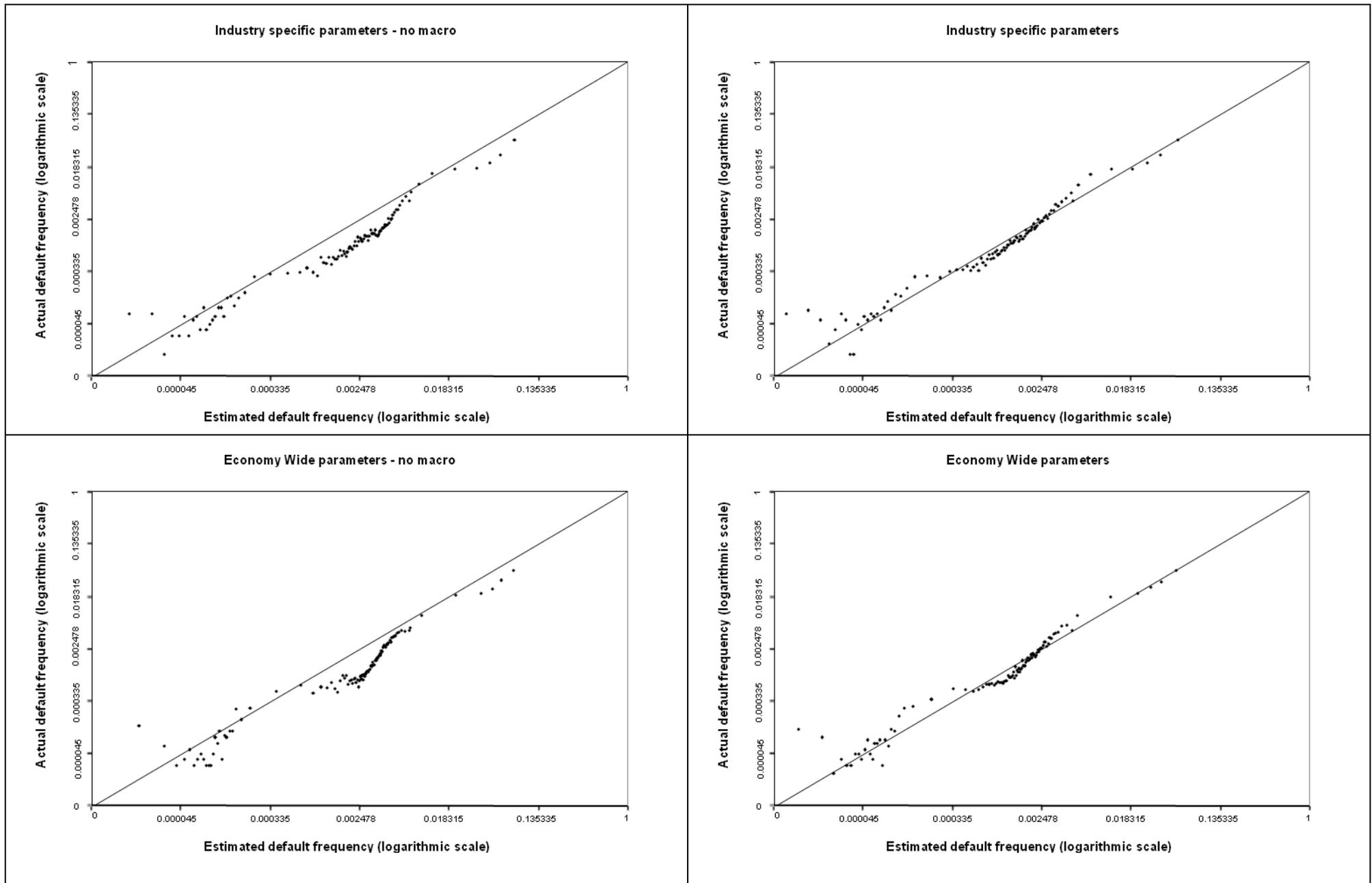


Figure 3: Sorted estimated default percentiles versus actual default frequencies for both economy-wide and industry-specific parameters. Left panel: Only firm-specific variables included (Table 1 models); Right panel: Macrovariables included (Table 2 models).

Appendix A Data

A.1 Definition of default

As described in Section 2, the default definition we adopt is the following: a firm is considered to be in default whenever one of the following events occurs: the firm is declared legally bankrupt; has suspended payments; has negotiated a debt composition settlement; is undergoing a reconstruction; or is distraint without assets. The data on Swedish public and private firms that we use to construct the default variable have been provided by Upplysningscentralen AB (UC), the main Swedish credit bureau, jointly owned by the Swedish banks. UC taps its information from Tingsrätten, the District Courts, Bolagsverket, the Swedish Companies Registration Office (SCRO), and Kronofogdemyndigheten, the Swedish Enforcement Authority, i.e. the institutions that deal with the legal formalities in firms' bankruptcy processes.^{A.1}

UC stores information on firms minor and major distress events in two databases, AM and JP. In the first database, AM, variables are constructed by giving each type of event a label AMTYPXX, a Swedish acronym for remark type, and an integer number suffix. For example, AMTYP12 is a dummy variable indicating if a firm has suspended its payments, or not. The second database, JP, contains 27 variables in total, on various milestones and stages for a broader category of major (mostly, but not exclusively) distress events that may occur for incorporated Swedish firms. From this database we take variables that are related to legal bankruptcy: "bankruptcy procedures started," "bankruptcy procedures concluded," "bankruptcy procedures concluded with a surplus," "bankruptcy procedures continued," and "declared bankrupt." Moreover, we also include: "negotiations on a debt composition settlement started," and "negotiations on a debt composition settlement concluded."

If any of the above distress-event dummy variables equals one at some point in our sample period, the firm in question is considered to be in default in that particular quarter. In the following quarter, we let the firm exit our data set. If more than one of these distress events are observed for a specific firm over our sample period, we assume the firm in question defaulted in the first instance. An additional variable we use from the second data set indicates if a "bankruptcy [was] cancelled" by a court. Over the whole sample period (i.e., in-sample and out-of-sample) this occurs 24 times, and 16 of these 24 events concern firms that subsequently end up in default. We treat firms for which the bankruptcy status was cancelled by the District

^{A.1} Currently, i.e., on July 9, 2010, there are 462 publicly listed firms out of roughly 250,000 active, limited liability firms.

Court as healthy until the data indicate otherwise. Moreover, we let firms that default but re-emerge from their default status exit the data set after the quarter in which default takes place; they re-enter in the quarter in which UC registered that the default status had been "removed."

Our decision to let firms that default exit the data set in the subsequent quarter is based on the following statistics: out of 161,550 defaults in the entire data set, 148,874 are terminal in the sense that subsequently no new information on the firms appears in any of the databases.^{A.2} The remaining 12,676 observations concern 6,638 firms that default twice within the sample period. Thus, 5,999 firms end up in terminal default at the second occurrence, while 339 re-emerge even after the second default. No firm defaults more than twice in the sample period.

Out of the 148,874 first-time-is-terminal defaults, an overwhelming majority of 145,684 are due to legal bankruptcy declarations. Roughly 60 percent of these firms experience a second default-triggering distress event simultaneously, i.e., in the same quarter. In almost all cases (98 percent) these are due to the event "bankruptcy proceedings initiated." In most (88 percent) of the remaining terminal defaults, i.e., those that are not legal bankruptcies, are associated with "distrain, no assets." The remaining distress events account for less than 1 percent of the first-time-is-terminal defaults.

For the firms that re-emerge after a default, the first default involves a legal bankruptcy in less than 2 percent of all cases and "distrain, no assets" in close to 90 percent. At their second default, these percentages are 97 percent due to legal bankruptcy, and 6 percent to "distrain, no assets," in the cases of terminal defaults. Among the firms that experience a second but non-terminal default, 62 percent has the status of the "distrain, no assets."

A.2 Balance sheet data

By means of a two-step procedure, the six financial ratios were selected as: earnings before interest, depreciation, taxes and amortization (EBITDA) over total assets (TA) (earnings ratio); interest payments (IP) over the sum of interest payments and earnings before interest, depreciation, taxes and amortization (interest coverage ratio); total liabilities (TL) over total assets (leverage ratio); total liabilities over total sales (TS) (debt ratio); liquid assets (LA) in relation to total liabilities (quick ratio); and inventories (I) over total sales (inventory turnover ratio).^{A.3}

^{A.2} Firms that are declared bankrupt at some point do not disappear from the databases that UC maintain. Firm identifiers (*organisationsnummer*) are unique and are never re-cycled by Swedish tax authorities.

^{A.3} It should be noted that the log-level of debt, in addition to the leverage ratio ($TL_{i,t}/TA_{i,t}$) for firm i in period t , contains predictive power for default. We therefore decided to include $TL_{i,t}$ as a separate variable, but scaled it with average total sales in period t to obtain a stationary ratio. Thus, the debt-to-sales ratio is defined as $\log(TL_{i,t}/TS_t)$, where TS_t denotes average total sales in period t .

Table A.1 in Appendix A.3 provides an account of the variables considered by other well-known studies in the literature.

First, the univariate relationship between a ratio and default risk was investigated. By visual inspection, ratios that are largely uncorrelated with default risk were eliminated from the set of candidate explanatory variables. Figure A.1 illustrates this for the six selected ratios by comparing default rates (jagged line) and the cumulative distributions of the variables (smooth line) for all observations in the period 1990Q1–2009Q2. The default rate for a given observation of a ratio is calculated as an average over the interval of ± 5000 adjacent observations in the empirical distribution of the ratio at hand. The cumulative distribution at any point X_0 on the X -axis gives the share of defaulted firms for which the financial ratio is smaller than X_0 . Given the density of the observations, there is a positive relationship between default and the leverage, interest coverage and turnover ratios, while there is a negative relationship for both the earnings and the liquidity ratios. Moreover, Figure A.1 suggests that the relationships between default and the earnings ratio, total liability over total sales ratio and interest costs over the sum of interest costs and earnings are all non-linear. For instance, for the interest coverage variable, this relationship is quite intuitive. The ratio turns highly negative if earnings are negative and slightly larger than interest payments in absolute value, which is associated with an increased risk of default. On the other tack, large interest payments and low earnings will also make the ratio large, positively, which is likewise associated with an increased default risk. Similar reasoning can be applied to the other ratios. In the second step in the selection procedure, variables that did not enter significantly were subsequently dropped one by one to get the final set of variables. For instance, standard variables like size (proxied by total sales) and age (proxied by the number of periods in the panel) were dropped in this second step as they were found to be insignificant in the full model.

Regarding the definition of the dummy variable for firms that have not submitted a financial statement, TTLFS, there are three points worth noting. First, this information is assumed to be available with a 6-quarter time lag, since financial statements for year τ are typically available in the third quarter of year $\tau + 1$. By letting the dummy variable equal unity with a 6-quarter time-lag, we account for the real-time delay. Second, given the way we define the population of existing firms, recently registered firms entering the panel would automatically be assigned $TTLFS = 1$ in the third quarter of their existence, since they have not, of course, issued any financial statement prior to entering. For these new firms, TTLFS has been set to 0 and the

accounting data variables have been taken from their first yearly balance sheet and income statement. Third, for defaulting firms that are in the panel but have on no occasion submitted an annual report, we also set TTLFS equal to 0. This is the case for about 20 percent of the 161,550 defaulting firms in the panel. So, although TTLFS turns out to be very important in the default-risk models, this feature is down-played rather than exaggerated in the construction of the variable.

A.3 Descriptive statistics for winsorized data

In Table A.2, we report the means and standard deviations for a set of accounting ratios, payment remarks, and a variable that measures the average elapsed time since the latest filing of a financial statement for the final data set that is used in the estimations in Section 3.

The table distinguishes between defaulted and non-defaulted firms, at the aggregate as well as the industry level, for the in-sample period 1990Q1 – 1999Q4, that is, the sub-sample period for which we will specify and estimate all subsequent models. For this period, we have a total of 8,106,176 observations of which 105,605 are defaults. Analyses of industry effects will be conducted at the one-digit level to ensure sufficiently many default observations in each industry in both the cross-sectional and the time series dimensions. The ten industries are; agriculture, manufacturing, construction, retail, hotel and restaurant, transportation, banking, finance and insurance, real-estate, consulting and rental, and finally a residual industry labeled “not classified”.

Because of the varying availability of data, the statistics in Table A.2 are calculated based on slightly different numbers of observations for the variables in a given industry. Dealing with microdata sets of this size invariably involves dealing with outliers. These observations would distort the estimation results if they were to be included in the logit model and therefore, we have winsorized the top and bottom 1 percent observations for the accounting variables in each industry.^{A.4} Given the large number of observations in our data set, this approach is practically more or less equivalent to simply deleting 1 percent of the observations that have accounting data that fall outside a certain region. Note that we choose to winsorize the observations in each industry separately, rather than at the aggregate level, thereby implicitly allowing for dispersion and different means in different industries. Table A.2 shows the descriptive statistics for the

^{A.4} Winsorization is quite common in the literature using financial ratios to avoid outliers that are created by near-zero denominators. Shumway (2001) winsorizes the top and bottom 1 percent of all observations. It should be emphasized that the results are robust to varying the winsorization rate between 0.5 and 2 percent.

winsorized microdata set.^{A.5}

A.4 Macro data

The aggregate time series are depicted in Figure A.2. They are: the output gap (i.e., the deviation of GDP from its trend value), the yearly inflation rate (measured as the fourth difference of the GDP deflator), the repo interest rate (a short-term nominal interest rate, set by the Riksbank), and the real exchange rate.

The repo rate was extremely high in the third quarter of 1992 due to the Riksbank having raised the so-called marginal interest rate to 500 percent, unexpectedly and temporarily, in an attempt to defend the fixed exchange rate. If the repo rate is not adjusted for this exceptional event, the estimation procedure would lead to underestimation of the importance of financial costs for default behavior. We therefore adjust the repo rate series in the third quarter of 1992 by means of a simple regression $R_t = b_1 + b_2 D923 + b_3 R_{t-1} + \varepsilon_t$. The estimated dummy coefficient \hat{b}_2 equals 28.2, and we therefore adjusted the repo rate 1992Q3 to equal 9.8 percent instead of 38 percent.

The output gap series is computed by HP-filtering GDP, where the smoothing coefficient λ is set to the standard value of 1,600. The real exchange rate is measured as the nominal TCW-weighted (TCW = trade competitive weights) exchange rate times the TCW-weighted foreign price level (CPI deflators) divided by the domestic CPI deflator. Note that a larger value for the real exchange rate implies a depreciation; hence a negative estimated coefficient for this variable implies that a depreciation on average reduces the risk of default at a given point in time. During the sample period, the real exchange rate is characterized by an upward trend (i.e., a tendency of gradual depreciation) and it is therefore detrended with the HP-filter as well to achieve stationarity. Appendix B.3 verifies that the results are robust with respect to our detrending procedure of the real exchange rate.

In the output-gap series in Figure A.2 there is clear evidence of the deep recession in the beginning of the 1990s with a negative output gap of more than 4 percent in 1993. The economic rebound of 1994-1995 is also evident, as is the IT-boom bust cycle in the late 1990. Finally,

^{A.5} Comparison of the descriptive statistics for unwinsorized data makes it clear that defaulted firms are disproportionately more affected when winsorizing all observations jointly. Since the PAYREMARK, TAXARREARS, PAYDIV and TTLFS are dummy variables that are unaffected by choice of winsorization procedure, a joint one could lead to underestimation of the importance of the accounting data variables in the default risk model relative to these dummy variables. To check the robustness of our chosen approach, we used an alternative approach where we truncated the healthy and defaulted firms separately. As expected, the estimation results of the default-risk model with this alternative winsorization suggest a somewhat larger role for the accounting ratios, but the overall picture remains the same.

Figure A.2 documents the exceptionally sharp downturn in the economy during late 2008 and the beginning of 2009.

Table A.1: Firm-specific explanatory variables used in papers focusing on the development of models of default risk

Paper	Year	Sample	Firms	Model	LHS variable	RHS variables					
						Liquidity	Profitability and efficiency	Solvency and leverage	Size	Other	
Altman	1968	66	Listed	Discriminant	Bankruptcy	(CA-CL)/TA	RE/TA, EBIT/TA, TS/TA	MVE/TC			
Frydman et al.	1985	200	Listed	Discriminant	Bankruptcy	LA/TA, CA/TA, LA/CL, CA/CL	CF/TD, EBIT/TA, NI / TA	MVE/TC, Ln(IC)	Ln(TA)		
Shumway	2001	39,745	Listed	Panel Logit	Bankruptcy	(CA-CL)/TA CA/CL	RE/TA, EBIT/TA, NI / TA, TS/TA	MVE/TL, TL/TA	Ln(MVE/market)	$r_E - r_M$ σ_E, σ_M	
Pesaran et al.	2006	[2,219]	Listed	Merton	Default		r_E	MVE TE/TA		σ_E CR	
Duffie et al.	2007	392,404	Listed	Hazard	Default		r_E			DtD	
Bonfim	2008	113,119	Mixed	Probit/Hazard	Loan default	LA/CL	NI/TA TS/TA	TL/TA, TE/TA		IR TSG	
Bharath, Shumway	2008	1.016m	Listed	Merton	Distance to default		NI/TA	MVE TD		σ_E	
""	"			Hazard	Time to default		NI/TA	MVE TD		PD	

Notes Table A.1: Lists of variables reflect the main model presented in each paper. In cases where no preferred model" was presented, the list reflects variables with significant variables in any model. The number of observations in a paper can vary dependent on model specification. CA = Current Assets, CL = Current Liabilities, TA = Total Assets, RE = Retained Earnings, EBIT = Earnings Before Interest and Taxes, TS = Total Sales, MVE = Market Value of Equity, TC = Total Credit, LA = Liquid Assets, CF = Cash Flow, TD = Total Debt, NI = Net Income, IC = Interest Coverage, r_E = Return on Equity, r_M = Market Return on Equity, σ_E = Volatility of Stock Returns, σ_M = Volatility of Market Stock Returns, TE = Total Equity, CR = credit Rating, DtD = Distance to Default, IR = Investment Rate, TSG = Total Sales Growth, PD = Probability of default. The term quick assets has been named liquid assets in this table because no consensus on the exact difference between liquid and quick assets. Some sources include inventories in liquid assets, but exclude them from quick assets.

Table A.2: Descriptive statistics for truncated firm-specific micro data 1990Q1-1999Q4

	Agriculture	Manufacturing	Construction	Retail	Hotel & Restaurant	Transport	Bank, Finance & Insurance	Real- Estate	Consulting & Rental	Not Classified	Total
Defaulted											
Number Obs	1643	13052	11843	32350	5457	5147	811	6470	14619	14213	105605
EBITDA/TA	-0.04 (0.35)	-0.01 (0.29)	0.01 (0.29)	-0.06 (0.35)	-0.12 (0.55)	0.01 (0.33)	-0.09 (0.66)	0.01 (0.27)	-0.03 (0.45)	-0.07 (0.45)	-0.04 (0.38)
TL/TA	1.09 (0.55)	0.97 (0.39)	0.96 (0.40)	1.04 (0.56)	1.20 (0.82)	1.01 (0.44)	1.12 (1.25)	1.02 (0.55)	0.94 (0.59)	0.99 (0.67)	1.01 (0.57)
LA/TL	0.14 (0.73)	0.13 (0.57)	0.18 (0.57)	0.17 (0.70)	0.20 (0.64)	0.19 (0.74)	0.43 (1.86)	0.18 (0.85)	0.34 (1.22)	0.36 (1.41)	0.22 (0.91)
I/TS	0.38 (0.84)	0.20 (0.31)	0.15 (0.42)	0.27 (0.45)	0.04 (0.06)	0.01 (0.04)	0.20 (1.33)	0.35 (1.54)	0.08 (0.25)	0.16 (0.48)	0.19 (0.56)
TL/TS	-1.12 (1.71)	-2.64 (1.78)	-1.64 (1.78)	-2.46 (1.69)	-1.72 (1.61)	-2.34 (1.70)	-2.56 (2.56)	-0.62 (2.19)	-1.69 (1.88)	-1.54 (2.18)	-1.98 (1.93)
IP/(IP+EBITDA)	0.25 (0.96)	0.24 (1.02)	0.18 (0.87)	0.23 (1.17)	0.19 (1.00)	0.23 (0.76)	0.23 (1.05)	0.41 (0.83)	0.15 (0.86)	0.20 (0.92)	0.22 (1.00)
PAYDIV	0.00 (0.03)	0.00 (0.05)	0.00 (0.05)	0.00 (0.04)	0.00 (0.04)	0.00 (0.03)	0.00 (0.04)	0.00 (0.04)	0.00 (0.05)	0.00 (0.03)	0.00 (0.04)
TTLFS	0.41 (0.49)	0.30 (0.46)	0.37 (0.48)	0.36 (0.48)	0.35 (0.48)	0.40 (0.49)	0.51 (0.50)	0.44 (0.50)	0.44 (0.50)	0.50 (0.50)	0.39 (0.49)
REMARK1	0.15 (0.36)	0.12 (0.32)	0.15 (0.36)	0.14 (0.35)	0.16 (0.37)	0.16 (0.37)	0.21 (0.41)	0.14 (0.35)	0.17 (0.37)	0.17 (0.37)	0.15 (0.36)
REMARK2	0.47 (0.50)	0.36 (0.48)	0.46 (0.50)	0.39 (0.49)	0.46 (0.50)	0.49 (0.50)	0.42 (0.49)	0.34 (0.47)	0.44 (0.50)	0.38 (0.49)	0.41 (0.49)
Non-defaulted											
Number Obs	259298	1229283	902494	2227966	258191	497011	120184	438139	1597081	470924	8000571
EBITDA/TA	0.14 (0.25)	0.12 (0.21)	0.12 (0.21)	0.08 (0.24)	0.08 (0.35)	0.17 (0.23)	0.08 (0.42)	0.09 (0.20)	0.14 (0.28)	0.13 (0.34)	0.11 (0.25)
TL/TA	0.70 (0.34)	0.69 (0.30)	0.71 (0.29)	0.74 (0.36)	0.87 (0.50)	0.72 (0.30)	0.68 (0.70)	0.80 (0.39)	0.63 (0.36)	0.65 (0.43)	0.71 (0.37)
LA/TL	0.48 (1.12)	0.43 (1.01)	0.44 (0.79)	0.43 (1.03)	0.38 (0.77)	0.46 (0.97)	1.42 (4.67)	0.42 (1.31)	0.90 (1.82)	0.99 (2.03)	0.58 (1.42)
I/TS	0.24 (0.57)	0.14 (0.24)	0.11 (0.31)	0.19 (0.34)	0.03 (0.06)	0.01 (0.05)	0.36 (1.80)	0.19 (1.11)	0.04 (0.18)	0.07 (0.34)	0.12 (0.44)
TL/TS	0.66 (1.30)	0.50 (1.85)	0.58 (1.81)	0.33 (0.90)	0.51 (1.25)	0.48 (1.74)	4.33 (26.14)	5.09 (20.15)	0.64 (2.26)	0.50 (1.58)	0.80 (6.03)
IP/(IP+EBITDA)	0.15 (0.71)	0.15 (0.72)	0.12 (0.72)	0.18 (0.87)	0.17 (0.79)	0.14 (0.54)	0.18 (0.94)	0.29 (0.68)	0.10 (0.75)	0.11 (0.75)	0.15 (0.77)
PAYDIV	0.14 (0.35)	0.15 (0.36)	0.13 (0.34)	0.13 (0.33)	0.06 (0.24)	0.13 (0.34)	0.13 (0.33)	0.08 (0.28)	0.16 (0.36)	0.12 (0.32)	0.13 (0.34)
TTLFS	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.04)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
REMARK1	0.00 (0.05)	0.00 (0.06)	0.00 (0.06)	0.00 (0.06)	0.01 (0.08)	0.00 (0.05)	0.00 (0.05)	0.00 (0.07)	0.00 (0.05)	0.00 (0.05)	0.00 (0.06)
REMARK2	0.03 (0.16)	0.03 (0.18)	0.04 (0.20)	0.03 (0.18)	0.06 (0.24)	0.04 (0.18)	0.02 (0.14)	0.03 (0.16)	0.03 (0.17)	0.03 (0.16)	0.03 (0.18)

Notes: The definition of variables are: EBITDA = earnings before taxes, interest payments and depreciations; TA = total assets; TL = total liabilities; LA = liquid assets; I = inventories; TS = total sales; IP = sum of net interest payments on debt and extra-ordinary net income; PAYDIV = a dummy variable equal 1 if the firm has paid out dividends during the accounting period and 0 otherwise; TTLFS = a dummy variable equal to 1 if the firm has not submitted an annual report in the previous year, and 0 otherwise; REMARK1 = a dummy variable taking the value of 1 if the firm has a payment remark due to one or more of the following events in the preceding four quarters; (i) a "non-performing loan" at a bank, or (ii) a bankruptcy petition, or (iii) issuance of a court order to pay a debt, or (iv) seizure of property; REMARK2 = a dummy variable taking the value of 1 if the firm is in various tax arrears.

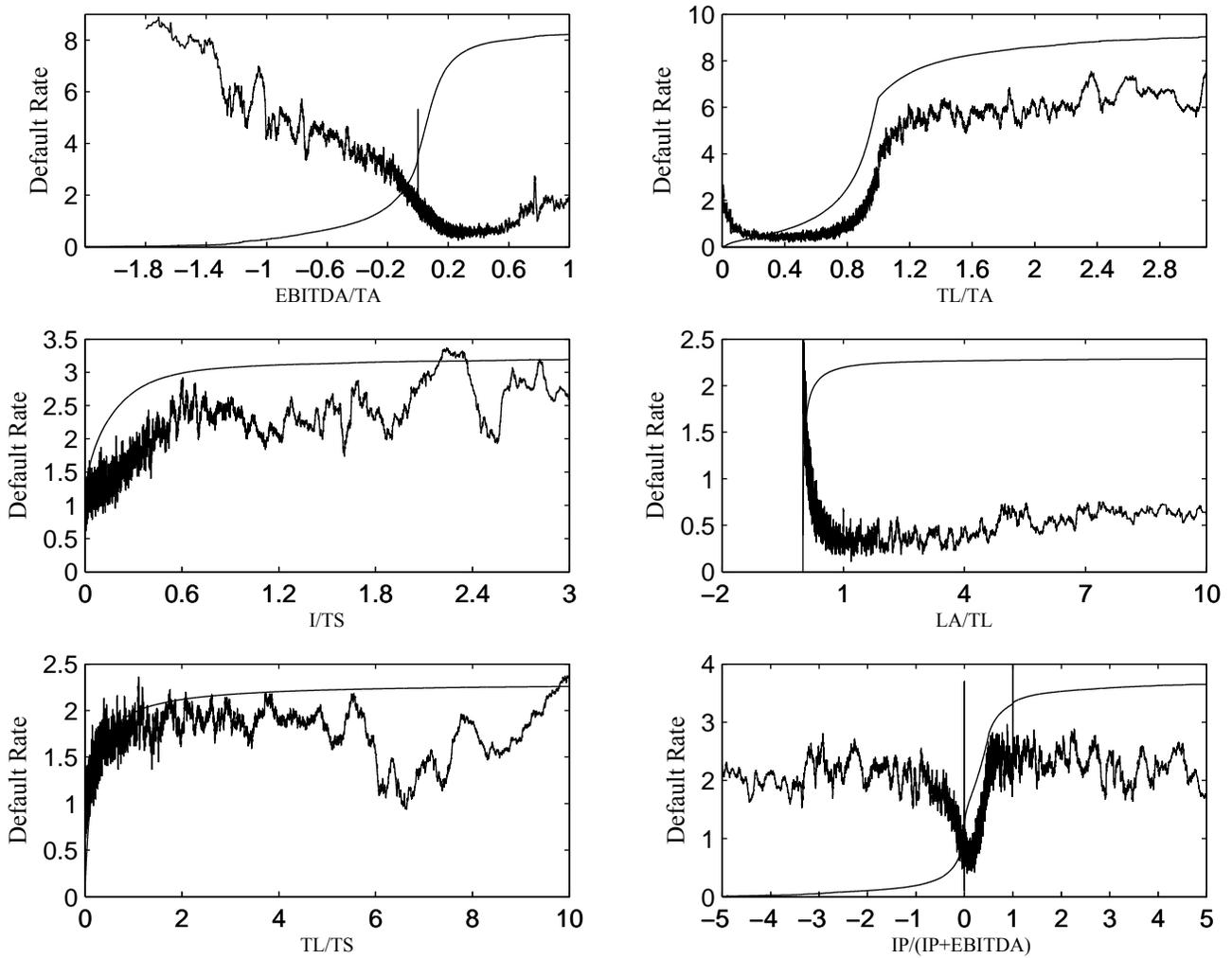


Figure A.1: Default rates and the cumulative distribution functions for the accounting data

The smooth lines are cumulative distributions of the default rate; the vertical axes on the right-hand-side have a minimum of zero and a maximum of one. Default frequencies at any point x_0 on the x-axis are computed as the average over 5,000 observations with values for the financial ratio smaller than x_0 and 5,000 observations greater than x_0 .

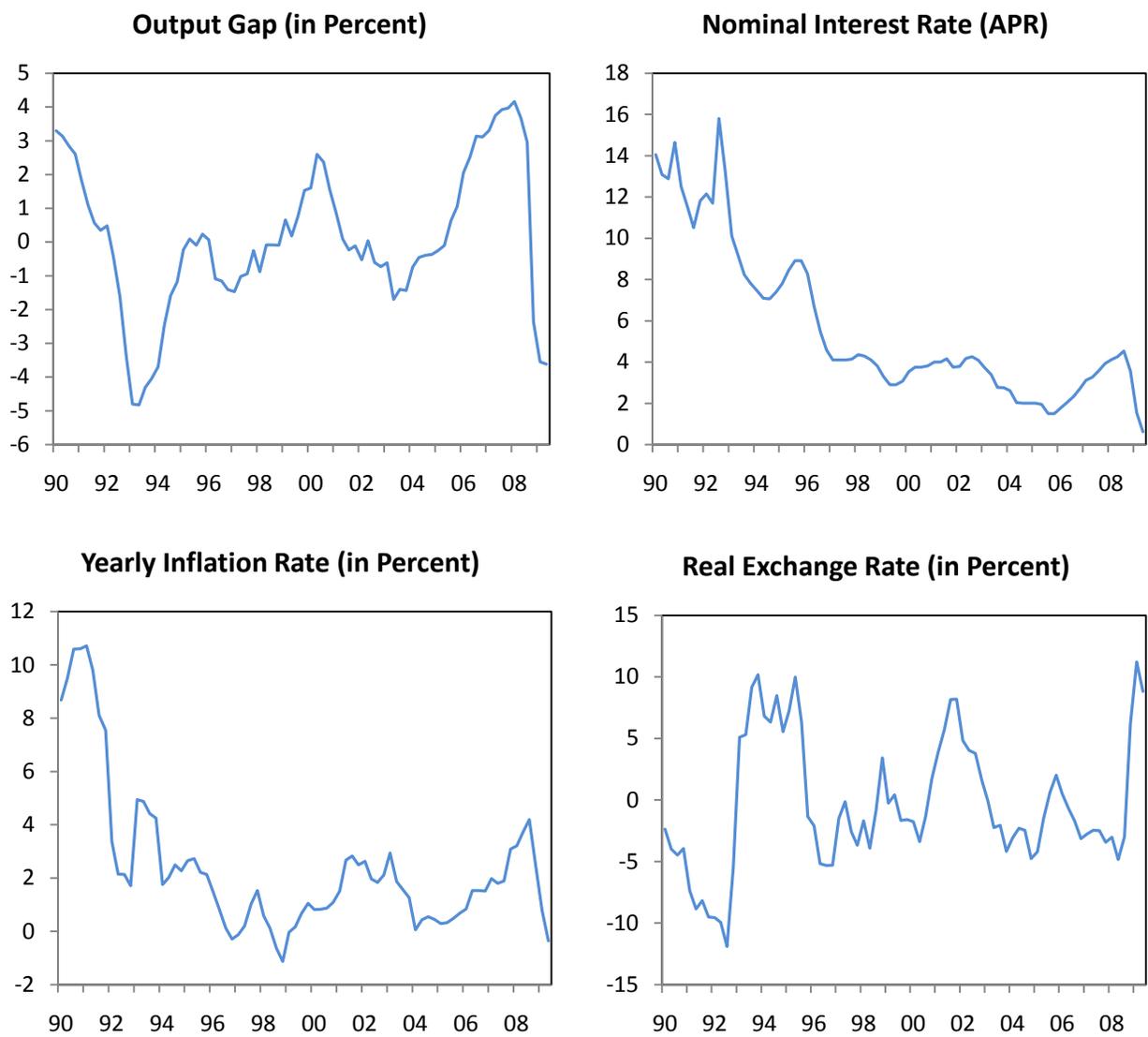


Figure A.2: Swedish Macro data 1990Q1-2009Q2 used in the estimated panel logit models

Appendix B Robustness analysis

The purpose of this appendix is to demonstrate that the results for the estimated models in Tables 1 and 2 are robust with respect to a number of perturbations. To keep the analysis tractable, we will restrict the analysis to the economy-wide models.

B.1 Role of remark data

Panel A in Table B.1 reports estimation results for the economy-wide models in Tables 1 and 2 when the PAYREMARK and TAXARREARS dummies have been dropped. For the sake of comparison, we also report the estimation results when the remarks variables are included in Panel A. As can be seen from comparing the results, the estimated coefficients for the accounting variables are similar irrespective of whether the remarks variables are included in the model or not. The coefficients for TL/TA and LA/TL increase when the remark variables are excluded, but the coefficient for EBITDA/TA is reduced. The remaining coefficients are roughly unaffected. Not in any case does exclusion of the Remark variables change the sign of the coefficients for the financial ratios. Thus, our estimation results for the financial ratios are not crucially affected by the inclusion of the remark variables. Nor do they diverge from the previous literature. However, it is clear that omitting the remark variables reduces the pseudo- R^2 measures, and thus reduces the ability of the model to rank the relative riskiness of firms. Turning to models where we include the macroeconomic variables in the regressions (c.f., Table 2), we again see that the coefficients are quite similar. An exception is the output gap coefficient, which turns out to be somewhat smaller in the model without remark variables. However, the coefficients for the output gap and the nominal interest rate are still of key importance, and thus the overall roles of macroeconomic variables are not affected by the presence of remark data. Once again, we find that extending the set of firm-specific factors with the remark variables allows us to increase the pseudo- R^2 measures substantially.

We conclude from this analysis that inclusion of remark data is not of key importance for the estimated impact of macroeconomic factors. Accordingly, our findings regarding the impact of macroeconomic factors would therefore be expected to hold in other countries, where payment remark data is not available. Nonetheless, the models' exceptional risk-ranking performance (as documented in Section 4.2) is clearly partly driven by the possibility to include remark data; without these data the out-of-sample risk-ranking performance would be worse than the in-

sample results reported by Shumway (2001) for publicly listed firms. But, given that the firms in our data set (i.e., the entire population of Swedish firms) are very heterogeneous, it is still remarkable that the risk-ranking performance of the models is quite acceptable even without the remark data.

B.2 Imputation of missing financial ratios

Panel B in Table B.1 reports estimation results for the economy-wide model in Tables 1 and 2 for regressions when only firms that have submitted complete financial statements for all sample periods are selected. In this case, obviously no imputation of missing financial ratios is carried out and the interaction dummy TTLFS is dropped since it will equal 0 for all firms. In addition, we drop all firms that have been active for a too short time period to meet the financial statement requirement. Comparing the results in Panel B with the benchmark results in Panel A, we see that the coefficients are little affected by our imputation procedure. The absolute values for the coefficients for EBITDA/TA and for LA/TL increase somewhat, while the remaining financial ratios are more or less the same. The coefficients for the remark variables, and for the DIVIDEND-variable are roughly the same, as are the coefficients for the macroeconomic variables. One interesting conclusion from this robustness analysis is that TTLFS is even more important than the remark data for the fit of the model at the firm level (pseudo- R^2). In the full benchmark model (III) in Panel A, the pseudo- R^2 is 0.34. In the model without remark data the pseudo- R^2 falls to 0.24. Excluding TTLFS (i.e., information on whether the firm issued a financial report in due time or not) leads to an even larger fall in pseudo- R^2 to about 0.17, suggesting that this indicator variable is the single most important predictor of default. This result has a lot of intuitive appeal, one would think that failure to compile a financial report should be a very serious signal that a firm is in the sort of trouble that could lead to a permanent exit.

B.3 Data frequency, sample period, and the real exchange rate

First we consider the effects of excluding the real exchange rate in the Table 2 model. In addition, we report results when the real exchange rate is calculated as the percentage deviation around a constant mean ($q_t = (Q_t - \bar{Q})/\bar{Q}$, where $\bar{Q} = \frac{1}{T} \sum_{t=1}^T Q_t$ for the period 1990Q1 – 2009Q2) instead of being HP-filtered. The results of these experiments are reported in Panel C in Table B.1. By and large, our results are not much affected by the choice of procedure for detrending

the real exchange rate. Differences in estimated coefficients for the financial ratios occur at the fourth decimal and are really miniscule. They are somewhat larger, though still small, for the set of indicator variables. For the other three macro variables we find that using the unfiltered real exchange rate reduces their coefficients marginally, while the real exchange rate coefficient is slightly increased. We conclude that using the filtered or unfiltered real exchange rate is of little consequence for the results in this paper. By a balanced regression argument we prefer the filtered real exchange rate as the benchmark, and we note in Section 2 that the results are robust w.r.t. to the detrending procedure.

In Panel D, we report results when we have estimated the model in Table 2 on an annual frequency instead of the quarterly frequency used in the paper. Again, we conclude that, on the whole, it is immaterial for the estimated parameters whichever choice of frequency is made. Noticeable exceptions are the coefficients for the debt ratio, TL/TA , and the output-gap, which both turn out stronger, while the coefficient for the real exchange rate is reduced, and that of inflation (imprecisely estimated) switches sign. Therefore, a quarterly model seems more appealing since it allows for more detailed forecasting and more interesting interpretations. It is nevertheless reassuring that our disaggregation procedure does not seem to introduce unwarranted biases.

Finally, to examine the stability of the coefficients for the Economy Wide models of Table 1 and 2, we have re-estimated them using the full sample period, 1990Q1 – 2009Q2, and present the results in Panel E in Table B.1. Comparing with the coefficients Panel E with those in Tables 1 and 2, we find only minor differences. The estimates for both the firm-specific and aggregate regressors are remarkably stable, consistent with the favorable out-of-sample results in Section 4 in both the cross-section and time series dimension.

B.4 Marginal effects

The explanatory variables in this paper have not been re-scaled to have the same mean, and therefore one cannot judge the importance of a particular variable from the size of its coefficient. The discussion about relative importance of explanatory variables in this paper is based on the relative sizes of the estimated t -statistics, since coefficient size is not sufficient for such inference. Alternatively, one can calculate the marginal contribution, or effect, from a variable at the mean, or the median, of the variable. Table B.2 report on such marginal effects and the calculations yield similar rankings of importance as the standard t -statistics, e.g., the output-gap and the

nominal interest rate are the more influential macroeconomic variables.

Table B.1: Sensitivity analysis for the economy-wide default risk models in Tables 1 and 2 1990Q1-1999Q4

	Panel A:				Panel B:				Panel C:			Panel D:		Panel E:		
	Robustness w.r.t. Remarks				Robustness w.r.t. TTLFS ^a				Robustness w.r.t. real ex. rate			Robustness w.r.t. freq		Robustness w.r.t. Sample		
	<i>Original sample (as in Tables 1 and 2)</i>				<i>Only include firms who have always submitted all FS data</i>				<i>Use alternative RER and omit RER</i>			<i>Quarterly vs. Annual Data</i>		<i>Full Sample (90Q1 - 09Q2)</i>		
	I	II	III	IV	I	II	III	IV	Benchmark	Alt. QD	Omit QD	Quarterly	Annual	I	II	
Micro Variables																
EBITDA/TA	-0,8371 <i>0,0106</i>	-0,7594 <i>0,0098</i>	-0,8221 <i>0,0107</i>	-0,7477 <i>0,0098</i>	-0,9929 <i>0,0125</i>	-0,9082 <i>0,0117</i>	-0,9747 <i>0,0125</i>	-0,8919 <i>0,0118</i>	-0,8221 <i>0,0107</i>	-0,8222 <i>0,0107</i>	-0,8259 <i>0,0107</i>	-0,8221 <i>0,0107</i>	-0,9411 <i>0,012</i>	-0,7139 <i>0,0080</i>	-0,7547 <i>0,0080</i>	
TL/TA	0,3948 <i>0,0071</i>	0,6982 <i>0,0063</i>	0,3832 <i>0,0072</i>	0,6976 <i>0,0064</i>	0,3968 <i>0,0083</i>	0,6926 <i>0,0074</i>	0,3857 <i>0,0084</i>	0,6977 <i>0,0075</i>	0,3832 <i>0,0072</i>	0,3834 <i>0,0072</i>	0,3752 <i>0,0072</i>	0,3832 <i>0,0072</i>	0,497 <i>0,008</i>	0,3200 <i>0,0053</i>	0,2982 <i>0,0054</i>	
LA/TL	-0,2067 <i>0,0061</i>	-0,3159 <i>0,0067</i>	-0,2045 <i>0,0061</i>	-0,3142 <i>0,0067</i>	-0,4073 <i>0,0121</i>	-0,6328 <i>0,0135</i>	-0,3915 <i>0,0121</i>	-0,6206 <i>0,0135</i>	-0,2045 <i>0,0061</i>	-0,2044 <i>0,0061</i>	-0,2060 <i>0,0061</i>	-0,2045 <i>0,0061</i>	-0,194 <i>0,006</i>	-0,1517 <i>0,0037</i>	-0,1425 <i>0,0037</i>	
I/TS	0,0573 <i>0,0052</i>	0,0682 <i>0,0046</i>	0,0444 <i>0,0053</i>	0,0607 <i>0,0046</i>	0,0622 <i>0,0059</i>	0,0718 <i>0,0052</i>	0,0490 <i>0,0060</i>	0,0633 <i>0,0052</i>	0,0444 <i>0,0053</i>	0,0442 <i>0,0053</i>	0,0427 <i>0,0053</i>	0,0444 <i>0,0053</i>	0,047 <i>0,006</i>	0,0875 <i>0,0043</i>	0,0655 <i>0,0045</i>	
TL/TS	0,1078 <i>0,0020</i>	0,0892 <i>0,0018</i>	0,1009 <i>0,0020</i>	0,0852 <i>0,0018</i>	0,0913 <i>0,0024</i>	0,0740 <i>0,0023</i>	0,0880 <i>0,0025</i>	0,0724 <i>0,0024</i>	0,1009 <i>0,0020</i>	0,1009 <i>0,0020</i>	0,1015 <i>0,0020</i>	0,1009 <i>0,0020</i>	0,104 <i>0,002</i>	0,0823 <i>0,0016</i>	0,0726 <i>0,0016</i>	
IP/(IP+EBITDA)	0,0655 <i>0,0040</i>	0,0737 <i>0,0039</i>	0,0558 <i>0,0039</i>	0,0674 <i>0,0038</i>	0,0595 <i>0,0046</i>	0,0638 <i>0,0046</i>	0,0500 <i>0,0044</i>	0,0567 <i>0,0045</i>	0,0558 <i>0,0039</i>	0,0557 <i>0,0039</i>	0,0553 <i>0,0039</i>	0,0558 <i>0,0039</i>	0,067 <i>0,005</i>	0,0651 <i>0,0036</i>	0,0515 <i>0,0035</i>	
PAYREMARK	1,7256 <i>0,0145</i>	1,8497 <i>0,0147</i>	1,8497 <i>0,0147</i>	1,834 <i>0,0147</i>	1,6866 <i>0,0165</i>	1,8043 <i>0,0168</i>	1,8043 <i>0,0168</i>	1,834 <i>0,0168</i>	1,8497 <i>0,0147</i>	1,8527 <i>0,0147</i>	1,8467 <i>0,0147</i>	1,8497 <i>0,0147</i>	1,834 <i>0,017</i>	1,8976 <i>0,0102</i>	1,9523 <i>0,0102</i>	
TAXARREARS	2,5654 <i>0,0092</i>	2,6839 <i>0,0094</i>	2,6839 <i>0,0094</i>	2,6839 <i>0,0094</i>	2,3323 <i>0,0108</i>	2,4713 <i>0,0111</i>	2,4713 <i>0,0111</i>	2,4713 <i>0,0111</i>	2,6839 <i>0,0094</i>	2,6857 <i>0,0094</i>	2,6593 <i>0,0094</i>	2,6839 <i>0,0094</i>	2,282 <i>0,010</i>	2,6088 <i>0,0070</i>	2,7241 <i>0,0071</i>	
Dividend	-3,1728 <i>0,0708</i>	-3,5374 <i>0,0706</i>	-3,0066 <i>0,0709</i>	-3,4573 <i>0,0706</i>	-3,1324 <i>0,0736</i>	-3,4098 <i>0,0736</i>	-2,9531 <i>0,0736</i>	-3,3082 <i>0,0736</i>	-3,0066 <i>0,0709</i>	-3,0046 <i>0,0709</i>	-3,0436 <i>0,0708</i>	-3,0066 <i>0,0709</i>	-3,221 <i>0,068</i>	-3,0896 <i>0,0456</i>	-2,9452 <i>0,0456</i>	
TTLFS	3,6937 <i>0,0084</i>	3,7990 <i>0,0074</i>	3,6076 <i>0,0085</i>	3,7532 <i>0,0075</i>	3,6937 <i>0,0084</i>	3,7990 <i>0,0074</i>	3,6076 <i>0,0085</i>	3,7532 <i>0,0075</i>	3,6076 <i>0,0085</i>	3,6133 <i>0,0085</i>	3,6164 <i>0,0085</i>	3,6076 <i>0,0085</i>	3,082 <i>0,010</i>	3,9342 <i>0,0073</i>	3,7527 <i>0,0074</i>	
Aggr. Variables																
Output gap			-0,1327 <i>0,0026</i>	-0,0782 <i>0,0024</i>			-0,1460 <i>0,0030</i>	-0,1061 <i>0,0029</i>	-0,1327 <i>0,0026</i>	-0,1205 <i>0,0025</i>	-0,1076 <i>0,0025</i>	-0,1327 <i>0,0026</i>	-0,210 <i>0,004</i>		-0,1119 <i>0,0018</i>	
Nominal interest rate			0,0731 <i>0,0016</i>	0,0501 <i>0,0015</i>			0,0594 <i>0,0018</i>	0,0351 <i>0,0018</i>	0,0731 <i>0,0016</i>	0,0522 <i>0,0020</i>	0,0971 <i>0,0014</i>	0,0731 <i>0,0016</i>	0,065 <i>0,002</i>		0,0703 <i>0,0012</i>	
GDP inflation			0,0116 <i>0,0023</i>	-0,0173 <i>0,0022</i>			0,0363 <i>0,0026</i>	0,0149 <i>0,0026</i>	0,0116 <i>0,0023</i>	0,0075 <i>0,0022</i>	-0,0231 <i>0,0020</i>	0,0116 <i>0,0023</i>	-0,061 <i>0,004</i>		0,0158 <i>0,0020</i>	
Real exchange rate			-0,0258 <i>0,0008</i>	-0,0071 <i>0,0003</i>			-0,0341 <i>0,0009</i>	-0,0198 <i>0,0009</i>	-0,0258 <i>0,0008</i>	-0,0272 <i>0,0008</i>		-0,0258 <i>0,0008</i>	-0,014 <i>0,001</i>		-0,0214 <i>0,0007</i>	
Mean log-likelihood	-0,046	-0,053	-0,046	-0,053	-0,039	-0,043	-0,039	-0,043	-0,046	-0,046	-0,046	-0,046	-0,125	-0,034	-0,034	
Pseudo R2	0,33	0,23	0,34	0,24	0,16	0,08	0,17	0,08	0,34	0,34	0,34	0,34	0,35	0,36	0,37	
Number of obs	8,106,176	8,106,176	8,106,176	8,106,176	7,949,015	7,949,015	7,949,015	7,949,015	8,106,176	8,106,176	8,106,176	8,106,176	2,207,382	16,928,521	16,928,521	

Notes: Coefficient estimates in bold style, standard errors in italics. The variables have not been scaled, so the importance of a variable cannot be interpreted directly from the size of the parameter estimate. The pseudo R² values have been calculated according to McFadden (1974). ^a The number of observations in these estimations are reduced by 157,161 as we are only including firms for which financial statement reports are available.

This means that we only include firms for which TTLFS = 0, and omit all defaulting firms that have never submitted financial accounting data. For reasons explained in more detail in Appendix A.2 in the paper, there are a large number of firms (27,492) that have never reported accounting data before they default, and these firms are all assigned TTLFS=0 in our analysis. Also these firms are excluded in the sub-sample considered here, so the number of defaults are only 64,189 in these estimations (as opposed to the 105,605 defaulting firms in Tables 1 and 2).

Table B.2: Marginal effects for the economy-wide default risk model

	Average of individual marginal effects				Marginal effects at the mean				Marginal effects at the median			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Micro Variables												
EBITDA/TA	-0,0083	-0,0084	-0,0081	-0,0083	-0,0030	-0,0037	-0,0020	-0,0025	-0,0037	-0,0063	-0,0023	-0,0040
TL/TA	0,0039	0,0077	0,0038	0,0077	0,0014	0,0034	0,0009	0,0024	0,0017	0,0058	0,0011	0,0038
LA/TL	-0,0020	-0,0035	-0,0020	-0,0035	-0,0007	-0,0015	-0,0005	-0,0011	-0,0009	-0,0026	-0,0006	-0,0017
I/TS	0,0006	0,0008	0,0004	0,0007	0,0002	0,0003	0,0001	0,0002	0,0003	0,0006	0,0001	0,0003
TL/TS	-2,2790	-2,2790	-2,2790	-2,2790	0,0004	0,0004	0,0003	0,0003	0,0005	0,0007	0,0003	0,0005
IP/(IP+EBITDA)	0,1507	0,1507	0,1507	0,1507	0,0002	0,0004	0,0001	0,0002	0,0003	0,0006	0,0002	0,0004
PAYREMARK	0,0171		0,0181		0,0062		0,0046		0,0076		0,0053	
TAXARREARS	0,0254		0,0263		0,0092		0,0067		0,0113		0,0076	
Dividend	-0,0314	-0,0392	-0,0295	-0,0382	-0,0113	-0,0170	-0,0075	-0,0117	-0,0139	-0,0293	-0,0085	-0,0187
TTLFS	0,0365	0,0421	0,0353	0,0415	0,0132	0,0183	0,0089	0,0128	0,0162	0,0314	0,0102	0,0203
Aggr. Variables												
Output gap			-0,0013	-0,0009			-0,0003	-0,0003			-0,0004	-0,0004
Nominal interest rate			0,0007	0,0006			0,0002	0,0002			0,0002	0,0003
GDP inflation			0,0001	-0,0002			0,0000	-0,0001			0,0000	-0,0001
Real exchange rate			-0,0003	-0,0001			-0,0001	0,0000			-0,0001	0,0000

Notes: Marginal effects for the explanatory variables in the economy-wide model. No standard errors are shown: statistical significance of the marginal effects is at the same level as for the parameter estimates in Table 3. Model III corresponds to the economy-wide model in Table 3, while the marginal effects in the columns marked I, II and IV refer to the models estimated for robustness purposes and displayed in Table B1.