

Validating Point-in-Time vs. Through-the-Cycle Credit Rating Systems

Manuel Mayer¹, Florian Resch², Stephan Sauer³

Preliminary version - please do not quote without permission

31 August 2017

Abstract

The validation of credit rating systems aims at identifying the quality of their credit risk estimates. Credit rating systems follow different rating philosophies, ranging from point-in-time (PIT) systems that reflect all currently available information to through-the-cycle (TTC) systems whose credit risk estimates are adjusted for cyclical changes in macroeconomic conditions. Although the terms PIT and TTC are widely used among credit rating agencies, banks, as well as supervisors, there is no consensus about their precise meaning.

This paper formalises a probabilistic framework to distinguish between credit scores, ratings and probabilities of default under PIT and TTC concepts to work out the key differences between PIT and TTC rating systems. We then analyze the validation of rating systems under both rating philosophies and conclude that the validation of TTC systems appears significantly more difficult than the validation of PIT systems. This finding can have important consequences for the optimal design of prudential requirements for ratings systems and financial institutions more generally, as credit rating systems play a key role in the context of micro- and macro-prudential supervision of the financial system.

¹Oesterreichische Nationalbank - Supervisory Statistics, Models and Credit Quality Assessment Division (email: manuel.mayer@oenb.at).

²Oesterreichische Nationalbank - Supervisory Statistics, Models and Credit Quality Assessment Division (email: florian.resch@oenb.at).

³European Central Bank - Risk Strategy Division (email: stephan.sauer@ecb.int). The views expressed in this paper are our own and do not necessarily represent the views of the European Central Bank, Oesterreichische Nationalbank or the Eurosystem.

1 Introduction

Financial market participants require accurate measures about an obligor's ability to fulfil his financial obligations in the future. Typically, the creditworthiness is characterised by a credit rating or credit score which is commonly associated with a probability of default (PD) which describes the likelihood that a firm experiences a credit event over a given time horizon. Credit events most commonly include criteria such as bankruptcy, payment delay, or unlikeliness to pay, however, the exact definition depends on the legal context. One key aspect according to which different approaches to estimating PDs can be classified is their point-in-time (PIT) vs. through-the-cycle (TTC) orientation.

Even though the terms PIT and TTC are widely used by practitioners, in academia and among regulators, a comprehensive literature review leads to the conclusion that a precise and generally accepted definition for these concepts appears to be non-existent at present. Unanimity only exists in the sense that PIT credit ratings make use of all information, both obligor-specific characteristics as well as overall macroeconomic conditions available at a certain point in time, whereas TTC credit ratings are adjusted for cyclical effects. Hence, the common understanding is that PIT ratings provide the most accurate and timely estimates of default probabilities, whereas TTC ratings provide a higher level of stability over time that comes at the cost of reduced timeliness and accuracy in predicting default events.

While this basic distinction between PIT and TTC systems is well established the type of cycle that is underlying TTC rating systems as well as the way it is measured varies considerably between TTC rating systems, as we will further discuss in Section 2 below. Moreover, many rating systems argue that they follow a hybrid approach that lies somewhere between the pure concepts of PIT and TTC. Other TTC rating systems do not explicitly define the cycle underlying their ratings but instead argue that the TTC nature of their rating systems implicitly stems from smoothing the explanatory variables of their rating models over time. The lack of a precise definition of TTC credit measures and the variety of methodological approaches towards the concept of TTC default probabilities that are found in practice make it difficult to interpret and compare ratings from different TTC rating sources. For the same reason, the validation of these rating systems constitutes a significant challenge as we will argue in more detail below.

In this paper we develop a simple but general model in which we provide formal definitions for both PIT and TTC credit scores, probabilities of default (PDs) and credit

ratings. Based on this model we characterize PIT and TTC credit risk measures and discuss the key differences between both rating philosophies. We then turn to the validation of default probabilities under both rating approaches and highlight that the validation of TTC rating systems is significantly more challenging than that of their PIT counterparts.

In particular, we stress that due to the lack of a precise definition of TTC default probabilities, TTC rating system have considerable leeway in choosing the type of cycle underlying their ratings, and how they measure it. Thus the validation of a TTC rating system involves assessing the economic validity of the cyclical factor. If the specific rating methodology of the TTC rating system is not available to the validator or the cyclical factor is taken into account implicitly, the underlying cycle of the rating system can be estimated from historical rating and default data based on the theoretical framework we present in this paper. This estimated cyclical factor can be compared to benchmark cyclical factors like real GDP growth.

Using the rating methodology of the TTC, we demonstrate how TTC ratings can be transformed into corresponding PIT ratings. The latter can then be validated using an assumption on the cyclical factor (e.g. the actual stance of the cycle or a stress scenario) and tested using the methodological toolkit available for validating PIT rating systems. This approach introduces significant model risk (and estimation risk if the parameters need to be derived from historical data) and thus the validity of the test results is diminished compared to validating a pure PIT rating system.

To sum up, we argue that compared to PIT rating systems, the validation of TTC rating systems is significantly more challenging from a methodological perspective as well as regards the amount of information that must be collected about the rating system. We further argue that the challenges of validating TTC rating systems together with the difficulties of interpreting and comparing ratings from TTC rating sources should be taken into account in regulatory decisions regarding the application of TTC rating systems.

This paper is structured as follows: In Section 2 we provide an overview of the current state of the literature about PIT and TTC default probabilities and rating methodologies. Section 3 establishes a probabilistic framework that allows a formal definition of PIT and TTC default probabilities and which we use to work out the key differences between the two rating concepts. In Section 4 we discuss the validation of both PIT and TTC rating systems. We highlight the challenges and limitations of validating TTC rating systems and demonstrate the approach using Standard & Poor's historical rating and default data. Section 5 summarizes and concludes the

paper.

2 Literature Overview

This section gives an overview of the current state of research of PIT and TTC credit measures in both academia as well as among credit rating agencies and supervisors. One point that we want to particularly highlight in this section is that despite the growing literature on TTC credit ratings there is still no consensus on the precise definition of a TTC credit rating except the general agreement that TTC ratings are adjusted for cyclical effects: the Basel Committee on Banking Supervision (2005) describes a PIT rating system as a rating system using all currently available obligor-specific and aggregate information to estimate an obligor's PD. On the contrary, a TTC rating system uses obligor-specific information but tends not to adjust ratings in response to changes in macroeconomic conditions. However, the types of these cyclical effects and how they are measured differ considerably in the literature as well as in practice.

First, a number of studies have come up with a formal definition of the concepts of PIT and TTC default probabilities and rating systems. These include Loeffler (2004) who explores the TTC methodology in a structural credit risk model based on Merton (1974) in which a firm's asset value is separated into a permanent and a cyclical component. In this model, building on Carey & Hrycay (2001), TTC credit ratings are based on forecasting the future asset value of a firm under a stress scenario for the cyclical component. Kiff et al. (2013) explore the TTC approach in a structural credit risk model in which the definition of TTC ratings follows the one applied by Loeffler (2004). They emphasize that "while anecdotal evidence from CRAs confirms their use of the TTC approach, it turns out that there is no single and simple definition of what TTC rating actually means". In contrast to the majority of studies in the literature that define PIT and TTC credit measures on the basis of a decomposition of credit risk into idiosyncratic and systematic risk factors, Hamilton et al. (2011) follow a frequency decomposition view in which a firm's credit measure is split up into a long-term credit quality trend and a cyclical component which are filtered from the firm's original credit measure by using a smoothing technique based on the Hodrick & Prescott (1981) filter. Furthermore, Hamilton et al. (2011) argue that in the existing literature there has been little discussion about whether the C in TTC refers to the business cycle or the credit cycle and highlight that these cycles differ considerably from each other regarding their length.

Aguais et al. (2008) describe a practical framework for banks for computing PIT and TTC PDs. They convert PIT PDs into TTC PDs based on sector-specific credit-cycle adjustments to distance-to-default credit measures of the Merton (1974) model derived from a credit rating agency’s rating or Moody’s KMV model.⁴ Furthermore, they qualitatively discuss key components of PIT-TTC default rating systems and how these systems can be implemented in banks. Carlehed & Petrov (2012) analyse PIT and TTC default probabilities of large credit portfolios in a Merton one-factor model. They define the TTC PD as the expected PIT PD, where the expectation is taken over all possible states of a systematic risk factor. Cesaroni (2015) proposes to translate PIT PDs into TTC PD by ex post smoothing the estimated PIT PDs with countercyclical scaling factors.⁵

Second, several studies analyse the ratings of major rating agencies as regards their PIT vs. TTC orientation. These include Altman & Rijken (2004) who find, based on credit scoring models, that major credit rating agencies pursue a long-term view when assigning ratings, putting less weight on short-term default indicators and hence indicating their TTC orientation. Loeffler (2005) shows for Standard & Poor’s and Moody’s rating data that these agencies have the policy to change a rating only if it is unlikely to be reversed in the future and argues that this can explain the empirical finding that rating changes lag changes of an obligor’s default risk, consistent with the general view of TTC ratings. Altman & Rijken (2006) analyse the TTC methodology of rating agencies from an investor’s PIT perspective and quantify the effects of this methodology on the objectives of rating stability, rating timeliness, and performance in predicting defaults. Among other results they find that TTC rating procedures delay migration in agency ratings, on average, by 1/2 year on the downgrade side and 3/4 year on the upgrade side and that from the perspective of an investor’s one-year horizon, TTC ratings significantly reduce the short-term predictive power for defaults. Several papers such as Amato & Furfine (2004) and Topp & Perl (2010) analyse actual rating data and show that these ratings vary with the business cycle, even though these ratings are supposed to be TTC according to credit rating agencies. Loeffler (2013) estimates long-run trends in market-based measures of one-year PDs using different filtering techniques. He

⁴Ingolfsson & Elvarsson (2010) follow a very similar approach to convert PIT PDs to TTC PDs by using the Kalman filter to estimate the credit cycle adjustment term from a bank’s historically incurred credit losses.

⁵Cesaroni (2015) uses the ratio of (i) the long-run average default rate or (ii) the maximum default rate to the current default rate as scaling factor. The first case leads to PDs similar to our TTC concept, whereas the second case corresponds to bottom-of-the-cycle PDs, which are sometimes referred to as TTC PDs.

shows that agency ratings contribute to the identification of these long-run trends, thus providing evidence that credit rating agencies follow to some extent a TTC concept. To summarize, many studies find that the ratings of major rating agencies show both PIT as well as TTC characteristics, which is consistent with the notion of hybrid rating systems.

Third, the rating philosophy is important from a regulatory and supervisory perspective as well as from an accounting perspective, not least because capital requirements for banks and insurance firms depend on credit risk measures. Studies that discuss TTC PDs in the context of Basel II or as a remedy for the potential pro-cyclical nature of Basel II include Repullo et al. (2010). Repullo et al. (2010) compare smoothing the input of the Basel II formula by using TTC PDs or smoothing its output with a multiplier based on GDP growth.⁶ They prefer the GDP growth multiplier because TTC PDs are worse in terms of simplicity, transparency, cost of implementation, and consistency with banks' risk pricing and risk management systems. Cyclicalities of credit risk measures also play an important role in the context of Basel III,⁷ which was implemented in the Capital Requirements Regulation (CRR) in the European Union.⁸ The CRR states that institutions shall have sound internal standards for situations where realised default rates deviate significantly from estimated PDs. According to the CRR, these standards shall take account of business cycles and similar systematic variability in default experience. In two separate consultation papers issued in 2016, the European Banking Authority (2016, p. 52-54) proposes to explicitly leave the selection of the rating philosophy to the banks, whereas the Basel Committee on Banking Supervision (2016, p.7) proposes to require banks to follow a TTC approach to reduce the variability in PDs and thus risk-weighted assets across banks.⁹

⁶Repullo et al. (2010) also discard alternative options to reduce the pro-cyclicality of Basel II, such as an auto-regressive adjustment rule and multipliers based on credit growth, stock-market prices, banks' profits or loan loss provisions.

⁷The Basel Committee on Banking Supervision (2011) introduced a series of measures to make banks more resilient to pro-cyclical effects. These measures have the objective to dampen excess cyclicalities of the minimum capital requirements, promote more forward looking provisions, conserve capital to build buffers that can be used in periods of stress, and achieve the broader macro-prudential goal of protecting the banking sector from periods of excess credit growth.

⁸Regulation (EU) No 575/2013 of the European Parliament and of the Council of 26 June 2013.

⁹The "variable scalar approach" to turn PIT PDs into TTC PDs was initially promoted in the United Kingdom for the implementation of Basel II, see Prudential Regulation Authority (2013) for a summary and Aguais et al. (2008) or Ingolfsson & Elvarsson (2010) for practical implementations. However, the PRA now expects banks to no longer use the variable scalar approach for residential mortgage portfolios because it has found that banks "are unable to distinguish sufficiently between

The adoption of the International Financial Reporting Standard (IFRS) 9 for the accounting of financial instruments as of January 2018 has further increased the relevance of rating philosophies. IFRS 9 requires an impairment allowance for financial assets held at amortised cost. The amount is based on the expected credit loss which shall explicitly be calculated with PIT PDs. IFRS 9 requires to calculate the expected credit loss for performing loans with PIT 12-month PDs; for underperforming and non-performing loans, the expected credit loss shall be calculated with PIT lifetime PDs for the expected life of the financial instrument. Building on Carlehed & Petrov (2012), Petrov & Rubtsov (2016) develop a methodology to first construct TTC rating grades and to then obtain PIT PDs needed for, e.g., IFRS 9. Skoglund (2017) is among the first papers to discuss the challenges to calibrate lifetime PIT PDs required by IFRS 9.

Finally, the rating philosophy should influence the validation of rating systems, but the challenges to validate TTC models have been largely ignored in the literature. Already Basel Committee on Banking Supervision (2005) stresses that in order to evaluate the accuracy of PDs reported by banks supervisors need to adapt their PD validation techniques to the specific types of banks' credit rating systems, in particular with respect to their PIT vs. TTC orientation. However, methods to validate rating systems have paid very little attention to the rating philosophy or focused on PIT models. For example, Cesaroni (2015) observes that predicted default rates are PIT and thus the validation of a rating system "should" operate on PIT PDs from a theoretical perspective. Petrov & Rubtsov (2016) explicitly mention that they have not yet developed a validation framework consistent with their PIT-TTC methodology. A key contribution of our paper is to address this important gap in the literature in a general and systematic manner.

3 Theoretical Framework

3.1 PIT vs. TTC Credit Risk Measures

In general, lenders face uncertainty about an obligor's ability to service his obligations in the future, i.e. lenders face credit risk. Against this background it is common practice to model the default of a firm as a stochastic event that is captured by a random indicator variable $Y_{i,\Delta t}$ which may take the following values:

movements in default rates that result from cyclical factors and those that result from non-cyclical reasons" (Prudential Regulation Authority, 2017, p. 17).

$$Y_{i,\Delta t} = \begin{cases} 1 & \text{if obligor } i \text{ defaults within the time period } (t; t + \Delta t), \\ 0 & \text{if obligor } i \text{ survives the period } (t; t + \Delta t). \end{cases} \quad (1)$$

The realization of the non-degenerated random variable $Y_{i,\Delta t}$ is governed by an ex ante default probability:

$$PD_{i,t} = P(Y_{i,\Delta t} = 1), \quad (2)$$

which can take values in the interval (0;1). Hence, at each point in time t and for any time horizon Δt a unique, true probability of default exists for each and every obligor i . This true PD is latent and hence non-observable. In credit risk measurement it is standard practice to model the true $PD_{i,t}$ on the basis of a true rating, or henceforth, credit score $S_{i,t}$ via a suitable link function ℓ :¹⁰

$$PD_{i,t} = \ell(S_{i,t}). \quad (3)$$

While there exists no agreement on the precise definition of TTC ratings in the literature (see Section 2), most studies define TTC ratings in the following general way: They decompose the credit risk of a firm into basic as well as cyclical risk factors. Based on this decomposition, they then define the TTC credit scores as the credit scores by setting the cyclical risk factors to their long run average.¹¹ In contrast, PIT scores are defined by taking into account both types of risk factors, basic as well as cyclical.

Hence, we model a firm's credit score $S_{i,t}$ as:

$$S_{i,t} = \alpha_i + X_{i,t} + \beta_i F_t, \quad (4)$$

where $X_{i,t}$ represents an idiosyncratic risk factor that captures a firm's business fundamentals such as its capital structure or management and F_t is a cyclical risk factor that represents aggregate information such as macroeconomic conditions.

For simplicity we will in the following assume that: $E(X_{i,t}) = 0$, $V(X_{i,t}) = \sigma_X^2$, $E(F_t) = 0$, $V(F_t) = \sigma_F^2$, and that $X_{i,t}$ and F_t are independent. The parameter β_i

¹⁰As discussed in Hornik et al. (2010) one of the most commonly used link functions is the standard normal cumulative distribution function Φ .

¹¹An alternative approach is to assume a stress scenario for the cyclical risk factors, leading to "bottom-of-the-cycle" credit risk measures.

captures the sensitivity of the credit score to the cyclical risk factor and α_i represents those parts of the basic and the cyclical risk factors that are constant over time. It follows that the expectation and variance of a firm's credit score are given by:

$$\begin{aligned} E(S_{i,t}) &= \alpha_i, \\ V(S_{i,t}) &= \sigma_X^2 + \beta_i^2 \sigma_F^2. \end{aligned} \tag{5}$$

We define the PIT and TTC credit scores of firm i as:

$$\begin{aligned} S_{i,t}^{PIT} &\equiv S_{i,t} = \alpha_i + X_{i,t} + \beta_i F_t, \\ S_{i,t}^{TTC} &\equiv \alpha_i + X_{i,t}. \end{aligned} \tag{6}$$

3.2 PIT vs. TTC Rating Systems

Consistent with Krahnert & Weber (2001) we define a rating system as a function:

$$R : \{\text{companies}\} \rightarrow \{\text{rating classes}\}. \tag{7}$$

This means that a rating system R assigns each element of a set of companies to a rating class, denoted for example by $\{A, B+, B, B-, \dots\}$. The assignment of companies to rating classes is based on the credit score and ensures that all companies within a rating class are reasonably homogeneous with respect to this credit score. Representing a certain (interval for the) credit score, each rating class is thus also associated with a corresponding (interval for the) probability of default (by means of the link function ℓ).

For simplicity we will in the following concentrate on perfect PIT and TTC rating systems, except the discussion of hybrid rating systems in Section 3.2.3. Hence, we will assume that the rating system is able to identify the true $S_{i,t}$ for each firm i . We will later relax this assumption by allowing for measurement errors when discussing the validation of PIT and TTC rating systems in Section (4).

3.2.1 PIT Rating Systems

In a PIT rating system a firm i is assigned to rating class c^{PIT} if its PIT credit score is equal to that underlying the rating class, denoted by s_c^{PIT} credit score of a PIT

rating class c^{PIT} in the following).¹² Hence, at each point in time t , the set of firms C_t^{PIT} assigned to the rating class c^{PIT} is given by:

$$C_t^{PIT} = \{i \mid S_{i,t} = s_c^{PIT}\}. \quad (8)$$

As a consequence, the default probability of each firm i assigned to the rating class c^{PIT} is given by:

$$PD_{i \in \{C_t^{PIT}\}, t} = \ell(s_c^{PIT}). \quad (9)$$

3.2.2 TTC Rating Systems

In contrast, in a TTC rating system a firm i is assigned to a rating class c^{TTC} on the basis of its TTC credit score. At each point in time t , the set of firms C_t^{TTC} assigned to the rating class c^{TTC} , with underlying TTC credit score s_c^{TTC} , is given by:

$$C_t^{TTC} = \{i \mid S_{i,t}^{TTC} = s_c^{TTC}\}. \quad (10)$$

Note that while in a PIT rating system, rating changes occur because of both, changes in the basic risk factor $X_{i,t}$ as well as changes in the cyclical factor F_t , in a TTC rating system rating changes occur only due to changes in the basic risk factor. Hence, a TTC rating system is expected to exhibit fewer rating changes than a PIT system.

It follows from equations (6) and (10) that the default probability of firm i that is assigned to rating class c^{TTC} is given by:

$$PD_{i \in \{C_t^{TTC}\}, t} = \ell(s_c^{TTC} + \beta_i F_t). \quad (11)$$

Whereas default probabilities are constant over time for a PIT rating class, they vary stochastically over time for a TTC rating class due to their dependence on the cyclical factor F_t . This result stems from the fact that a TTC rating system ignores the cyclical part of a firm's credit score when assigning it to a rating class. Furthermore, note that even though in a TTC rating system firms are assigned to

¹²Without affecting our main results we assume that all firms within a rating class are not only reasonably homogenous but identical with respect to their credit scores.

rating classes on the basis of their TTC credit scores, their observed default processes are determined by the true default probability $PD_{i,t}$ (see Section 3.3).

3.2.3 Hybrid Rating Systems

In practice many rating system providers argue that they follow a hybrid approach (see, e.g., European Banking Authority (2013, p. 28)). In order to formally introduce a hybrid rating system and differentiate it from PIT and TTC systems we define the hybrid credit score of firm i at time t as:

$$S_{i,t}^{HYB} \equiv \alpha_i + X_{i,t} + \delta\beta_i F_t, \quad (12)$$

where $\delta \in [0; 1]$ denotes the degree to which a hybrid rating system follows a PIT approach. In a hybrid rating system a firm i is assigned to a rating class c^{HYB} on the basis of its hybrid credit score. At each point in time t , the set of firms C_t^{HYB} assigned to the rating class c^{HYB} , with underlying hybrid credit score s_c^{HYB} , is given by:

$$C_t^{HYB} = \{i \mid S_{i,t}^{HYB} = s_c^{HYB}\}. \quad (13)$$

Hence, the default probability of a firm i that is assigned to the rating class c^{HYB} is given by:

$$PD_{i \in \{C_t^{HYB}\}, t} = \ell(s_c^{HYB} + (1 - \delta)\beta_i F_t). \quad (14)$$

Hence, a hybrid rating system represents a compromise between a PIT and TTC rating system in the sense that only a certain fraction of the cyclical component of a firm's credit score is accounted for when it is assigned to a rating class. As a consequence, as for TTC rating systems the default probability of a rating class of a hybrid rating system varies stochastically over time.

This section has shown that hybrid rating systems can be easily included in our framework. For the sake of simplicity, we do not further discuss hybrid systems below. The challenges to validate TTC systems highlighted in Section 4 become less relevant the closer δ gets to 1, i.e. a pure PIT system.

3.3 Default Frequencies in PIT vs. TTC Rating Systems

In this section we explore the characteristics of default frequencies in PIT and TTC rating systems. In the following we let n_t^c be the number of firms assigned to either a PIT rating class c^{PIT} or a TTC rating class c^{TTC} . Furthermore, we denote the default frequency of either a PIT or TTC rating class by $D_t^c = \frac{\sum_{i \in \{C_t\}=1}^{n_t^c} Y_{i,\Delta t}}{n_t^c}$ and a realization of the default frequency by d_t^c . The corresponding default patterns for T periods and M rating classes are denoted by $\mathbf{D} = [D_t^c]_{T \times M}$ and $\mathbf{d} = [d_t^c]_{T \times M}$.

3.3.1 Default Frequencies in PIT Rating Systems

In a PIT rating system, according to equation (9) the default probability of a firm i assigned to a PIT rating class c^{PIT} is given by $PD_{i \in \{C_t^{PIT}\}, t} = \ell(s_c^{PIT})$ and hence the number of defaults of a PIT rating class, $n_t^c D_t^c$ follows a binomial distribution:

$$P(n_t^c D_t^c = n_t^c d_t^c) = \binom{n_t^c}{n_t^c d_t^c} \ell(s_c^{PIT})^{n_t^c d_t^c} (1 - \ell(s_c^{PIT}))^{n_t^c - n_t^c d_t^c}. \quad (15)$$

3.3.2 Default Frequencies in TTC Rating Systems

We next turn to TTC rating systems. According to equation (11) the default probability of a firm i assigned to a TTC rating class c^{TTC} is given by:

$$\widetilde{PD}(F_t) \equiv PD_{i \in \{C_t^{TTC}\}, t} = \ell(s_c^{TTC} + \beta_{c^{TTC}} F_t), \quad (16)$$

where we assumed for simplicity that all firms in the rating class c^{TTC} have the same sensitivity $\beta_{i \in \{C_t^{TTC}\}} = \beta_{c^{TTC}}$ to the cyclical factor and we use the shorthand notation $\widetilde{PD}(F_t)$ for the true PD of a firm assigned to a TTC rating class c^{TTC} .

We define the PIT and TTC default probabilities of firm i assigned to a TTC rating class c^{TTC} as:¹³

¹³Note that alternatively the TTC default probability of a firm could be defined on the basis of its TTC score: $PD_{i \in \{C_t^{TTC}\}}^{TTC, alt} = \ell(\mathbb{E}(s_c^{TTC} + \beta_{c^{TTC}} F_t)) = \ell(s_c^{TTC})$ which, due to Jensen's inequality, will in general differ from the expression given in (17). We choose the former definition since, as shown below, this ensures that the TTC default probability of a firm assigned to a TTC rating class equals the expected default rate in that rating class.

$$\begin{aligned}
PD_{i \in \{C_t^{TTC}\}, t}^{PIT} &\equiv \widetilde{PD}(F_t), \\
PD_{i \in \{C_t^{TTC}\}, t}^{TTC} &\equiv \overline{PD} \equiv E(PD_{i \in \{C_t^{TTC}\}, t}^{PIT}) = E(\widetilde{PD}(F_t)).
\end{aligned} \tag{17}$$

Conditional on a realization of the cyclical factor $F_t = f_t$ the random indicator variables $Y_{1, \Delta t}, \dots, Y_{n_t, \Delta t}$ are independent and each have default probability $\widetilde{PD}(f_t)$. We can think of a realization of the default frequency $D_t^c = d_t^c$ as being generated in two steps. First, a cyclical factor $F_t = f_t$ and hence a default probability $\widetilde{PD}(f_t)$ is realized. Second, given the default probability $\widetilde{PD}(f_t)$ the number of defaults $n_t^c d_t^c$ is drawn from a binomial distribution with n_t^c trials. Conditional on a realization of the cyclical factor $F_t = f_t$, the probability of observing a number of defaults $n_t^c D_t^c = n_t^c d_t^c$ is given by:

$$P(n_t^c D_t^c = n_t^c d_t^c \mid F_t = f_t) = \binom{n_t^c}{n_t^c d_t^c} \widetilde{PD}(f_t)^{n_t^c d_t^c} (1 - \widetilde{PD}(f_t))^{n_t^c - n_t^c d_t^c} \tag{18}$$

The unconditional distribution is then obtained by integrating over the distribution of F_t :

$$\begin{aligned}
P(n_t^c D_t^c = n_t^c d_t^c) &= \int_{-\infty}^{\infty} P(n_t^c D_t^c = n_t^c d_t^c \mid F_t = f_t) g(f_t) df_t \\
&= \int_{-\infty}^{\infty} \binom{n_t^c}{n_t^c d_t^c} \widetilde{PD}(f_t)^{n_t^c d_t^c} (1 - \widetilde{PD}(f_t))^{n_t^c - n_t^c d_t^c} g(f_t) df_t,
\end{aligned} \tag{19}$$

where g denotes the density of F_t . It can be shown that the unconditional expectation and variance of the default frequency D_t^c are given by:¹⁴

$$\begin{aligned}
E(D_t^c) &= \overline{PD} \\
V(D_t^c) &= \frac{\overline{PD}(1 - \overline{PD})}{n_t^c} + \frac{(n_t^c - 1)}{n_t^c} V(\widetilde{PD}(F_t)).
\end{aligned} \tag{20}$$

The unconditional expectation of the default frequency for both the PIT PD and the TTC PD assigned to firm i is equal to the TTC default probability $PD_{i \in \{C_t^{TTC}\}}^{TTC}$ following directly from the definition in equation (17). The unconditional variance of the default frequency equals the binomial variance around the TTC PD, i.e. the

¹⁴We derive equation (20) in Section 6.1 in the appendix.

expected PIT PD of the TTC class, plus the variance of the PIT PD of the TTC class triggered only by changes in the cyclical factor F_t .

Table 1 summarises some stylized facts about the differences between PIT and TTC rating systems. In Section 4 we will address these deviations when we discuss methods validating both types of rating systems, focusing on the key differences of relevance for this paper.

Table 1: Summary of differences between PIT and TTC rating systems

Feature	PIT model	TTC model
<i>Key differences of relevance for this paper</i>		
Use of information	Reflects all available information	Abstracts from / removes information on (business) cycle
Effect of credit quality changes caused by (business) cycle	Rating migration	None (TTC rating constant through the cycle)
Number of rating changes	Many (caused by idiosyncratic and systematic factors)	Few (caused only by idiosyncratic factors)
True PD of a rating class	Constant	varies with F_t , negatively correlated with the (business) cycle
Variance of realised default rates	Low	High
Observed default rates by rating grade / PD bucket over (business) cycle	Constant (same as long-run average default rate)	Varying: increasing in downturn, decreasing in upswing; equal to long-run average default rate only in the middle of cycle
Default correlation between obligors	(Close to) 0	Positive
<i>Other differences implicitly included in the model in this paper</i>		
Capital requirements over (business) cycle under Basel II (Basel Committee on Banking Supervision, 2005, see, e.g.)	Varying: increasing in downturn, decreasing in upswing	Constant (note: in practice, portfolio adjustments will lead to changes)
PD under stressed macroeconomic conditions (Basel Committee on Banking Supervision, 2005, see)	Positively correlated with the (business) cycle	Constant

4 Validation of PIT and TTC Rating Systems

So far it has been assumed that a rating system is able to perfectly allocate firms to rating classes based on their true PIT and TTC scores. In practice, however, the assignment of obligors to rating classes is subject to measurement error and hence there is the need to perform ex-post validation of credit rating systems.

Several approaches have been proposed for validating rating systems (for an overview, see Basel Committee on Banking Supervision, 2005). The most widespread approach, which is referred to as back-testing, is to compare ex-post realised default rates with ex-ante estimates of probabilities of default. Other methods include the assessment of discriminatory power (see, e.g., Lingo & Winkler, 2008) or benchmarking where ratings from different sources are compared (see, e.g., Hornik et al., 2007). Based on the theoretical foundations defined in Section 3, we will in the following provide a framework for back-testing PIT and TTC rating systems.

4.1 Validation of PIT Rating Systems

In the following let for a PIT rating class c^{PIT} , with PD PD_c^{PIT} , the estimated PIT credit scores be given by $\widehat{S}_{i,t}$. Hence, the set of firms assigned to a PIT rating class is now given by:

$$C_t^{PIT} = \{i \mid \widehat{S}_{i,t} = \widehat{\alpha}_i + \widehat{X}_{i,t} + \widehat{\beta}_{cTTC} \widehat{F}_t = s_c^{PIT}\}, \quad (21)$$

and the corresponding estimated default probabilities of firms in rating class c^{PIT} are given by: $\widehat{PD}_{i \in \{C_t^{PIT}\},t}^{PIT} = \ell(\widehat{S}_{i,t}) = \ell(s_c^{PIT})$. In the validation of credit rating systems it is standard practice to apply the assumption of Section 3.2, namely that firms within a rating class are not only homogeneous but identical with respect to their credit scores and default probabilities, and to assess the calibration quality of a rating class by comparing its underlying default probability with the corresponding observed default frequency.

The most basic and common validation method involves testing whether the rating system is not underestimating probabilities of default of rating class c using a one-sided test of the form:

$$H_0^c : PD_c^{PIT} \leq \widehat{PD}_c^{PIT} \quad \text{vs.} \quad H_1^c : PD_c^{PIT} > \widehat{PD}_c^{PIT}, \quad (22)$$

where PD_c^{PIT} and \widehat{PD}_c^{PIT} denote the true and estimated PIT PD of obligors assigned to rating class c . Alternatively, one may test for both under- and overestimation of PDs by employing a two-sided statistical test:

$$H_0 : PD_c^{PIT} = \widehat{PD}_c^{PIT} \quad \text{vs.} \quad H_1^c : PD_c^{PIT} \neq \widehat{PD}_c^{PIT}, \quad (23)$$

Under the null hypothesis, the observed one-year default frequency for rating class c follows a binomial distribution defined in equation (15). Hence, in the case of a PIT rating system the null hypothesis is given by (22) or (23) and the calibration quality can be tested by using standard tests for binomial distributions.¹⁵ For a one-sided alternative the binomial test is the uniformly most powerful test, for two-sided alternatives an overview of different approaches is given in Aussenegg et al. (2011).

4.2 Validation of TTC Rating Systems

Recall from Section 3.2.2 that in a TTC rating system, firms are assigned to a TTC rating class c^{TTC} based on their TTC credit scores $S_{i,t}^{TTC}$. Hence, the set of firms assigned to a TTC rating class is given by:

$$C_t^{TTC} = \{i \mid \widehat{S}_{i,t}^{TTC} = \widehat{\alpha}_i + \widehat{X}_{i,t} = s_c^{TTC}\}, \quad (24)$$

As a consequence, while all firms in a TTC rating class share the same TTC credit score (abstracting from measurement error at this point), their actual (PIT) credit scores, their true (PIT) PDs, and hence the number of defaults observed in each TTC rating class, will vary over time. This variation is due to the variation of the cyclic factor F_t and is ultimately the reason why validating a TTC rating system is more challenging than validating a PIT rating system as described above. Validating a TTC rating system thus requires to assess both, the idiosyncratic component as well as the cyclic factor of the rating system. To this end, the validator will make use of all observable data which may indicate a miscalibration: the realized default rates per TTC rating class and the realizations of the cyclical factor.

Starting with the cyclical factor we find that by choosing a specific type of cyclical factor (e.g. business cycle or credit cycle) and the way how it is measured, the

¹⁵Note that we assume here that the validator knows the PDs $\ell(s_c^{PIT})$ for each rating class. If they are not known they can be estimated as outlined in section (4.2).

TTC rating system implicitly chooses the amount by which the true (PIT) PDs, and hence the number of observed defaults, will vary over time in each TTC rating class. As discussed in Sections 1 and 2 there is no consensus on the exact definition of a TTC rating system and hence there is also no consensus on the definition and measurement of the cyclical factor F_t . Thus, in practice, TTC rating systems have considerable leeway in choosing the type of cyclical factor they employ and hence they have considerable leeway in choosing the amount of variation they allow in their true (PIT) PDs over time.

For the validation of the cyclical risk component of a TTC rating system this means that the validator checks whether the choice of the cyclical factor is in line with the overall objectives and application of the TTC rating system. More specifically, she might check how the cyclical factor chosen by the TTC rating system correlates with standard macroeconomic variables such as real GDP growth, output gap, or aggregate lending. Alternatively, the validator might compare the cyclical factor of the TTC rating system with a benchmark cyclical factor (we will discuss this possibility in more detail below).

The validation of a TTC rating system crucially depends on the amount of information that is available to the validator. In this paper we will investigate two boundary cases. In the first case the validator has full information about the model parameters and the cyclical factor chosen by the TTC rating system. In the second and more realistic case, the validator observes only the number of defaults and the number of obligors in each TTC rating class without any knowledge of the underlying model parameters and the cyclical factor.

4.2.1 Validation of a TTC Rating System with Full Information

After checking the suitability and plausibility of the cyclical factor employed by the TTC rating system the validator can validate the TTC rating system by computing the corresponding PIT PDs for each rating class of the TTC system according to equation (16) and then test these PIT PDs against observed default rates, in the same way as outlined in section (4.1). For example, assume that the validator knows that for a TTC rating class with 1000 rated entities c^{TTC} : $s_c^{TTC} = -2.5$, $\beta_c^{TTC} = 0.45$, $\sigma_F = 0.45$, $F_t = -1$, and that $\ell = \Phi$. The validator can then compute the corresponding PIT default probability for this TTC rating class as according to equation 16:

$$PD_{i \in C_t^{TTC}}^{PIT, TTC} = \Phi(-2.5 - 0.45^2) = 0.0034 \quad (25)$$

Given a PIT PD of 34 bp, a standard hypothesis test, for example as represented in equation (22), can be performed. For a significance level of 5%, the critical value for rejecting the null hypothesis that the rating class is well calibrated amounts to 7. Hence, observing 7 or more defaults for this rating class would result in rejection of the null hypothesis. Note, however, that the model parameters as well as the cyclical factor will typically not be known to the validator and hence this approach is likely to be of limited relevance for validators in practice.

In practice, however, most validators do not have full information about the model parameters and the cyclical factor chosen by the TTC rating system. Moreover, many rating systems do not explicitly estimate a cyclical factor but instead argue that their TTC (or hybrid) nature is due to the fact that they smooth their explanatory variables over the business (or credit) cycle. In this case the approach described above is not feasible. The validator can, however, infer the model parameters and the cyclical factor of the TTC rating system from observed default rates. We will turn to this possibility in the next section.

4.2.2 Validation of a TTC Rating System with Limited Information

If the validator has no information about the underlying model parameters and the cyclical factor of a TTC rating system she can estimate the relevant model parameters from observed default rates. In the following we show how a TTC rating system can be validated based on estimating its idiosyncratic and cyclical risk component in line with the theoretical framework presented in this paper. By using Standard & Poor’s historical rating and default data we estimate the implied cyclical factor which can then be compared to a benchmark cyclical factor (e.g. real GDP growth) or be used as a benchmark factor for the validation of the cyclical components of other TTC rating systems¹⁶ As we will further discuss below, estimating the idiosyncratic and cyclical risk components requires a sufficient time series of data which in practice will be a significant obstacle for following this approach.

Our data covers the time period ranging from 1981 to 2016. Thus, our dataset

¹⁶It is important to note that the TTC nature of Standard & Poor’s ratings has been challenged in the literature, see Section 2. As a consequence, the difference between PIT PDs at the peak or bottom of the cycle and a “true” TTC PD would be even greater than based on Standard & Poor’s data.

ultimately comprises $T = 36$ years. For each year we are equipped with the one-year default rate for the respective grade for corporate entities including financial institutions on a rating grade level. We pool information on corporates and financials and further aggregate numbers over the distinct modifiers of the respective major rating categories to finally obtain data on a rating class level. Since default rates for rating classes AAA as well as AA have been largely zero we decide to focus only on rating classes A, BBB, BB, B, and CCC/C. Since, the number of obligors for each rating class are not published by Standard & Poor's we set these numbers to A: 1000, BBB: 1000, BB: 500, B: 500, and CCC/C 100, which roughly reflects the typical pool observed in the literature.

A summary of the dataset is given in table (4) in the appendix. In line with several other rating systems, Standard & Poor's does not publish explicit credit scores (s_c^{TTC}) (or equivalently TTC PDs PD^{TTC}) for its different rating classes.

Combining equations (16) and (18) the log-likelihood of observing a default pattern \mathbf{D} and obligor pattern $\mathbf{n} = [n_t^c]_{T \times M}$ is given by:

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{f} \mid \mathbf{n}, \mathbf{d}) = \sum_{t=1}^T \sum_{c=1}^M \log \binom{n_t^c}{n_t^c d_t^c} + \sum_{t=1}^T \log H_t \quad (26)$$

$$H_t = (\Phi(\mu_c + \sigma_c f_t))^{(n_t^c d_t^c)} \cdot (1 - \Phi(\mu_c + \sigma_c f_t))^{(n_t^c - n_t^c d_t^c)},$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_M)'$, and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_M)'$, $\mathbf{f} = (f_1, \dots, f_T)'$, $\mu_c = s_c^{TTC}$, and $\sigma_c = \beta_c^{TTC} \sigma_F$. Furthermore, as outlined in section 3.3, $\mathbf{d} = [d_t^c]_{T \times M}$ denotes the observed default pattern. It is clear from the specification above that the mean of \mathbf{f} : $f_m = \frac{1}{T} \sum_{t=1}^T f_t$ and the mean of $\boldsymbol{\mu}$: $\mu_m = \frac{1}{M} \sum_{c=1}^M \mu_c$ cannot be separately identified. In the same way the variance $\text{Var}(f_t)$ and the mean level of $\boldsymbol{\sigma}$: $\sigma_m = \frac{1}{M} \sum_{c=1}^M \sigma_c$ cannot be distinguished.

In order to identify the parameters given in (26) we impose two further restrictions: $\widehat{f}_m = 0$ and $\text{Var}(f_t) = 1$, i.e. we estimate the normalized implied cyclical factor.¹⁷ Maximizing the log-likelihood subject to these restrictions yields parameter estimates $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \dots, \widehat{\mu}_M)'$, and $\widehat{\boldsymbol{\sigma}} = (\widehat{\sigma}_1, \dots, \widehat{\sigma}_M)'$, $\widehat{\mathbf{f}} = (\widehat{f}_1, \dots, \widehat{f}_T)'$ as well as the corresponding standard deviations of the estimated parameters (based on the Fisher Information Matrix) denoted by $sd_{\widehat{\mu}_c}$, $sd_{\widehat{\sigma}_c}$, and $sd_{\widehat{f}_t}$ which are summarized in Table (2).

¹⁷Note that the restriction $\widehat{f}_m = 0$ can be justified by observing a sufficiently long time series.

As expected the parameters $\mu_c = s_{cTTC}$ monotonically increase over the different rating classes, implying that the TTC PD increases when going from rating class “A” to “CCC/C”. In contrast, the sensitivity parameter $\sigma_c = \beta_{cTTC}$ is fairly stable over the different rating classes and amounts to 0.30 on average. As mentioned above Standard & Poor’s does not publish explicit credit scores (s_c^{TTC}) (or equivalently TTC PDs PD^{TTC} for its different rating classes. In the case where these credit scores s_c^{TTC} are available¹⁸, they can be tested against the estimated parameters $\hat{\mu}_c$, which capture the idiosyncratic risk components estimated by the TTC rating system. This can be done for example by a Wald test: under the null hypothesis, the test statistics $\frac{\hat{\mu}_c - x}{sd_{\hat{\mu}_c}^2}$ asymptotically follows a Chi-squared distribution with one degree of freedom.

The estimated cyclical factor $\hat{\mathbf{f}} = (\hat{f}_1, \dots, \hat{f}_T)'$ is presented in figure (1) and shows a clear cyclical pattern over the sample period. Moreover, comparing the estimated cyclical factor to US real GDP growth¹⁹ for the same sample period, as shown in figure (1), reveals that the estimated cyclical factor closely tracks US GDP growth over most of the sample period²⁰.

Table (3) shows the average estimated PIT PDs: $\frac{1}{T} \sum_{t=1}^T \widehat{PD}_{i \in \{C_t^{TTC}\}, t}^{PIT} = \frac{1}{T} \sum_{t=1}^T \Phi(\hat{\mu}_c + \hat{\sigma}_c \hat{f}_t)$, the average observed default frequencies $d^c = \frac{1}{T} \sum_{t=1}^T d_t^c$, as well as the 95% empirical quantile of the estimated PIT PDs $q_{95} PD_c^{TTC, PIT}$ for each rating class. The table reveals that the estimated PIT PDs closely track the observed default frequencies in each rating class. Figures (2) and (3) show the estimated PIT PDs $\widehat{PD}_{i \in \{C_t^{TTC}\}, t}^{PIT} = \Phi(\hat{\mu}_c + \hat{\sigma}_c \hat{f}_t)$ (indicated by “estimated”) and observed defaults rates (indicated by “observed”) for the different rating classes over time. We observe that implicit factor allows to closely track the realised default rates of all rating classes.

¹⁸If the rating system reports TTC PDs for each rating class, the associated TTC credit scores s_c^{TTC} have to be computed based on equation (17) and the estimated standard deviations of the cyclical risk components σ_c .

¹⁹US real GDP growth was taken from the IMF World Economic Outlook 2016 database.

²⁰The plain correlation coefficients between the two time series amounts to -0.26. Note that US real GDP growth was normalized ($X^{normalized} = \frac{X - \mu}{\sigma}$, where X denotes real GDP growth, μ denotes the mean of real GDP growth and σ denotes its standard deviation) and that in figure (1) it was multiplied by (-1) for better comparability.

Table 2: ML Estimates – Implied Cyclical Factor

ML estimates for historical Standard and Poor’s default data, 1981 - 2016

	A	BBB	BB	B	CCC/C
$\hat{\mu}_c$	-3.35	-3.10	-2.43	-1.78	-0.74
$sd_{\hat{\mu}_c}$	0.08	0.06	0.03	0.02	0.02
$\hat{\sigma}_c$	0.27	0.39	0.25	0.30	0.29
$sd_{\hat{\sigma}_c}$	0.07	0.05	0.03	0.02	0.02
	mean	st.dev.	min	max	
\hat{f}_t	0.00	1.00	-1.43	1.97	
$sd_{\hat{f}_t}$	0.08	0.00	0.07	0.08	

Table 3: ML Estimates – Estimated PIT PDs

Average estimated PIT PDs and observed default frequencies (in %)

	A	BBB	BB	B	CCC/C
$PD_c^{TTC,PIT}$	0.06	0.20	0.93	4.43	23.94
d^c	0.06	0.21	0.93	4.43	23.86
$q_{95} PD_c^{TTC,PIT}$	0.20	0.81	2.34	10.55	40.95

Having estimated the implied cyclical factor, the validation of the TTC rating system then comes down to validating the implied cyclical factor as discussed above. Hence, the validator might compare the estimated cyclical factor with a benchmark cyclical factor or she might compare it with a set of macroeconomic variables, as indicated in Figure (1). However, when the validator has to estimate the cyclical factor, the validation analysis is now subject to a number of additional limitations. First, estimation of the cyclical factor requires a sufficient amount of historical data. Second, the estimation requires the assumption of a model framework. Third, the estimated cyclical factor will contain possible measurement error in the cyclical factor.

In our example, US real GDP growth may be a key cyclical indicator for the sample of rated global corporates, in particular because for most years the majority of the ratings and the observed defaults in the sample refer to US firms. However, it may be questionable whether our model fully reflects the influence of the cyclical factor on S&P ratings, whether US real GDP growth is the right cyclical factor, and

whether the differences between our estimated cyclical factor and the US real GDP growth in Figure (1) reflect measurement error or should lead to the rejection of the hypothesis of a well-calibrated TTC rating model. In the next section we will compare the validation of PIT and TTC rating systems and discuss the practical implications of the challenges of validating TTC rating systems.

4.3 Comparison of the approaches

Already the Basel Committee on Banking Supervision (2005) stressed that in order to evaluate the accuracy of PDs reported by banks, supervisors need to adapt their PD validation techniques to the specific types of banks' credit rating systems, in particular with respect to their PIT vs. TTC orientation. This paper highlights that validating TTC models by back testing their calibration quality is very challenging.²¹

We find that PIT systems are relatively easy to validate and require as a minimum only the time horizon of the PD (usually one year) as data.²² The testing procedures are based on a small set of assumptions (e.g. independence of defaults) which can be easily relaxed if necessary. Results are available in timely manner and any significant deviations from the ex-ante PDs are highly reliable and may therefore be associated with concrete policy actions.

In view of the discussion of differences between PIT and TTC systems, it is most important that defaults are usually correlated because of negative realisations of systematic risk.²³ In the words of the Basel Committee on Banking Supervision (2005,

²¹Mayer & Sauer (2017) explain why tests of discriminatory power are not sufficient to validate TTC models: first, standard tests of discriminatory power are portfolio-dependent and the results thus cannot be used for the comparison of models applied to different portfolios. Second, if all firms have the same sensitivity β_{cTTC} to the systematic factor, two rating systems will rank all debtors equally if their only difference is that one system follows a PIT approach and the other one is a TTC system. Both systems will have the same discriminatory power, but the TTC system may still suffer from poor calibration quality. Third, if banks use TTC models for capital requirements purposes, the interests of banks and supervisors are likely to be aligned for identifying bad debtors, i.e. discriminatory power, whereas this is less clear for calibration quality. Banks can benefit from lower capital requirements if their models systematically underestimate the level of PDs, while they can systematically correct for this underestimation in their business decisions such as their average interest rates.

²²We ignore here the statistical problems associated with small sample sizes that are particularly pronounced for low-default portfolios, which are not uniquely defined but historically include exposures to sovereigns, banks, insurance companies or highly rated large corporates.

²³An example of default correlation potentially independent of systematic risk is the joint default of several companies belonging to the same group. Such default correlation based on non-systematic

p. 51), unconditional defaults are not independent, but defaults are independent conditional on a realisation of the systematic factor. The concrete value for default correlation is unknown.²⁴ Since the systematic factor F_t captures all systematic aspects, it can be argued that the relevance of default correlation for the validation of PIT models is significantly reduced or even absent, thus justifying the assumption of independent defaults in validation tests.²⁵

While TTC systems deliberately correct for the cyclical effects in the final rating outcome, default events are governed by the true but latent PIT realisations. As highlighted in the previous section, PIT PDs exceed their corresponding TTC PDs by a factor of around three at the 95%-quantile (see Table 3). The greater variance in the realised default rate of TTC systems impacts the validation of calibration quality because it implies wider confidence bands around TTC PDs than PIT PDs. Hence, for any given observed difference between the PD and the realised default rate, the likelihood that this difference indicates that the model is *not* well-calibrated is lower for a TTC model than for a PIT model. In other words, it is much easier to detect a poorly-calibrated PIT model that underestimates PDs than an equally poorly-calibrated TTC model if the statistical test takes into account the different confidence bands associated with the different rating philosophies; the poor calibration of the TTC model can then only be detected around the bottom of the cycle. In addition, estimating confidence bands for TTC rating systems requires a sufficiently large sample of historical default data and a set of model assumptions which are very difficult to justify without knowledge of the exact rating model and the specification of the systematic factor used by the TTC system.

Since the publication of Basel Committee on Banking Supervision (2005), some additional tests of calibration quality have been developed (see, e.g., Coppens et al., 2007, 2016). However, also these more recent tests do not address the issue of default correlation and thus do not properly reflect the nature of TTC models. The challenge to validate the calibration of TTC models has also been explicitly highlighted by several credit rating agencies claiming to follow a TTC rating approach towards their supervisor (see European Securities and Markets Authority, 2015).²⁶ Notably, the

risk should be irrelevant for the back-testing of credit risk models if the sample size is large enough.

²⁴In the Basel framework, default correlation is modelled via asset correlation in the Basel II one factor model. The asset correlation is in the range between 0.12 and 0.24 in the Basel formula for risk-weighted assets under the internal ratings-based approach, depending on PD, maturity of loan etc.

²⁵The (close to) zero (asset) correlation for PIT models is formalised and estimated in Rösch (2005), for example (see also, e.g., Blümke, 2011).

²⁶According to European Securities and Markets Authority (2015, p. 16), “ESMA has observed

European Securities and Markets Authority (2015) argues that credit rating agencies should overcome this challenge given that credit ratings are used not only for the appropriate rank ordering, but also e.g. for regulatory purposes in the context of the standardised approach for banks' or insurance firms' capital requirements according to Basel III and Solvency II. In the view of the European Securities and Markets Authority (2015), it would raise standards in the industry if CRAs consistently use a minimum standard of statistical measures in demonstrating the predictive power of their methodologies.

The Basel Committee on Banking Supervision (2005) concluded that statistical tests alone cannot be sufficient to adequately validate a rating system and need to be complemented by qualitative assessments. Hence, any application of a statistical technique has to be supplemented by qualitative checks and banking supervisors conduct extensive analysis of banks' internal models under Pillar 2 of the Basel framework.²⁷ However, qualitative assessments are almost by definition more art than science, and thus potentially diverging interests between banks and their supervisors become particularly relevant. The interests in having good credit risk estimates for capital requirements are of course much better aligned between the bank and the supervisor if the bank uses the IRB output not only for capital requirements purposes, but also for internal purposes, such as the bank's internal risk management and reporting, credit decisions and pricing of loans. Hence, a key element of the qualitative assessment according to the Basel framework is the "use test": banks must actually use the IRB output for all purposes; deviations are only allowed if they are reasonably explained to the supervisor.²⁸

that the majority of the credit rating agencies find assessing the predictive power of their methodologies challenging. In certain cases, credit rating agencies state that their ratings are based on an ordinal rather than a cardinal ranking which limits the extent to which internal expectations are relevant to the validation of the predictive power of a methodology, given the volatility of these expectations across the economic cycle."

²⁷The Basel framework puts the primary responsibility for validation of IRB models on the bank (see, e.g., Art. 185 of the CRR). Basel Pillar 2 requires that all banks make their own assessments of capital required, including risks not properly captured in Pillar 1 (minimum capital requirements) by IRB models. In addition, banking supervisors have the possibility for applying additional capital charges as the consequence of a variety of tools, including stress tests, under the "Supervisory Review and Evaluation Process" (SREP) (see, e.g., European Banking Authority, 2014). Such stress tests usually use stressed (PIT) PDs, which can also be considered as bottom-of-the-cycle PDs.

²⁸Art. 179 of the CRR requires that "Where institutions use different estimates for the calculation of risk weights and for internal purposes, it shall be documented and be reasonable." Other qualitative aspects that supervisors consider include the model design, the data quality and availability and governance aspects such as the independence of the rating process (see, e.g., Deutsche

The challenges to validate TTC models can have important consequences regarding their use for regulatory purposes, both from a micro- and a macro-prudential perspective. Mayer & Sauer (2017) discuss different alternatives to TTC models to avoid the potential pro-cyclicality of PIT-based capital requirements.

5 Conclusion

This paper analyses the differences between PIT and TTC credit risk measures. A comprehensive literature review has revealed that there is no consensus about the precise meaning of the two different rating concepts at present. Assembling the unambiguous notions from the literature, we have built a formal probabilistic framework comprising precise definitions for the PIT and the TTC concept. We use this framework to theoretically analyse the key differences and links between TTC and PIT PDs, scores and ratings and the corresponding realised default rates.

Furthermore, the literature review shows that there is very limited research on the validation of TTC rating systems, in contrast to the extensive literature on the validation of PIT rating systems. We show first that TTC rating systems cannot be validated in the same way as PIT systems. We then explain how TTC rating systems can be validated and highlight the methodological challenges involved. We provide a concrete example for the validation of a TTC system using historical Standard and Poor's rating and default data. The significant challenges involved in the validation of TTC rating systems are amplified in practice due the hybrid nature of many TTC systems as well as the huge variety of TTC methodologies. These challenges should be taken into account in the current discussion about the use of TTC rating systems for regulatory purposes.

Bundesbank, 2003).

6 Appendix

6.1 Default Frequencies

In this section we derive further characteristics of the default frequency D_t in a TTC rating system. To simplify notation we write $\widetilde{PD}(F_t) = \widetilde{PD}_t$. By conditioning we find that:

$$\begin{aligned} \mathbb{E}(Y_{i,\Delta t}) &= \mathbb{E}(\mathbb{E}(Y_{i,\Delta t} | \widetilde{PD}_t)) = \overline{PD}, \\ \mathbb{V}(Y_{i,\Delta t}) &= \mathbb{E}(Y_{i,\Delta t}^2) - (\mathbb{E}(Y_{i,\Delta t}))^2 = \overline{PD} - \overline{PD}^2 \\ &= \overline{PD}(1 - \overline{PD}). \end{aligned} \tag{27}$$

The covariance of the default indicator variables $Y_{i,\Delta t}$ is given by:

$$\begin{aligned} \text{Cov}(Y_{i,\Delta t}, Y_{j,\Delta t}) &= \mathbb{E}((Y_{i,\Delta t} - \mathbb{E}(Y_{i,\Delta t}))(Y_{j,\Delta t} - \mathbb{E}(Y_{j,\Delta t}))), \\ &= \mathbb{E}(\mathbb{E}(Y_{i,\Delta t} Y_{j,\Delta t} | \widetilde{PD}_t)) - \overline{PD}^2 \\ &= \mathbb{E}(\widetilde{PD}_t^2) - \overline{PD}^2 \\ &= \mathbb{V}(\widetilde{PD}_t). \end{aligned} \tag{28}$$

From these results we can derive the expectation and variance of the default frequency D_t as:

$$\begin{aligned} \mathbb{E}(D_t) &= \frac{\sum_{i=1}^{n_t} Y_{i,\Delta t}}{n_t} = \overline{PD} \\ \mathbb{V}(D_t) &= \frac{1}{n_t^2} \left(\sum_{i=1}^{n_t} \mathbb{V}(Y_{i,\Delta t}) + 2 \sum_{i=1}^{n_t-1} \sum_{j=i+1}^{n_t} \text{Cov}(Y_{i,\Delta t}, Y_{j,\Delta t}) \right) \\ &= \frac{\overline{PD}(1 - \overline{PD})}{n_t} + \frac{(n_t - 1)}{n_t} \mathbb{V}(\widetilde{PD}_t). \end{aligned} \tag{29}$$

Note that when the number of firms n_t becomes large, the variance of the default frequency converges to that of \widetilde{PD}_t :

$$\mathbb{V}(D_t) \rightarrow \mathbb{V}(\widetilde{PD}_t) \quad \text{as } n_t \rightarrow \infty. \tag{30}$$

Next consider the average default frequency $AD = \frac{1}{T} \sum_{t=1}^T D_t$ over a time period of length T . By the Central Limit Theorem we know that for large T , AD is approximately normally distributed with:

$$\begin{aligned}
 E(AD) &= \frac{\sum_{t=1}^T E(D_t)}{T} = \overline{PD}, \\
 V(AD) &= \frac{\sum_{t=1}^T V(D_t)}{T^2} \\
 &= \frac{1}{T^2} \sum_{t=1}^T \left[\frac{\overline{PD}(1 - \overline{PD})}{n_t} + \frac{(n_t - 1)}{n_t} V(\widetilde{PD}_t) \right].
 \end{aligned} \tag{31}$$

6.2 Tables

Table 4: Descriptive Statistics - Default Frequencies

Default Frequencies (%), 1981 - 2016

	A	BBB	BB	B	CCC/C
Mean	0.06	0.21	0.93	4.43	23.86
Min	0.00	0.00	0.00	0.2	0.00
Max	0.40	1.00	4.20	13.80	49.00

This table shows the average, minimum, and maximum of Standard and Poor's historical default rates over the period from 1981 to 2016 as described in Section 4.2.

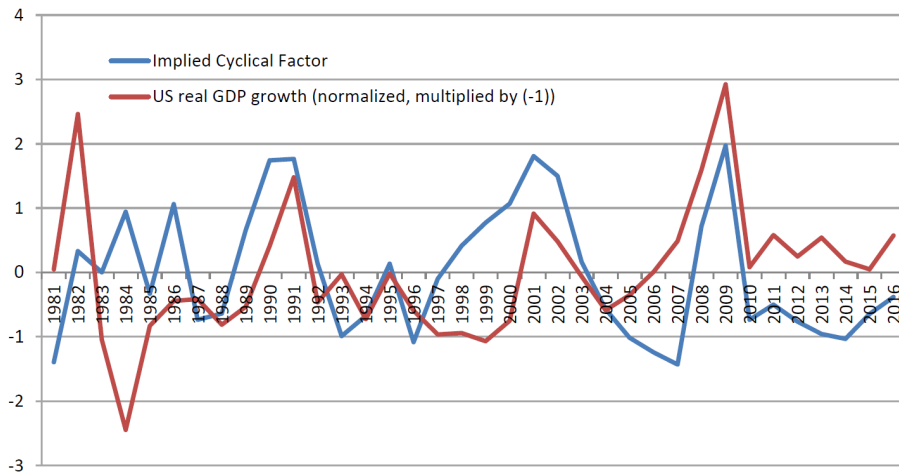


Figure 1: Implied Cyclical Factor and US real GDP growth (normalized). US real GDP growth was normalized ($X^{normalized} = \frac{X-\mu}{\sigma}$, where X denotes real gdp growth, μ denotes the mean of real gdp growth and σ denotes its standard deviation) and multiplied by (-1) for better comparability.

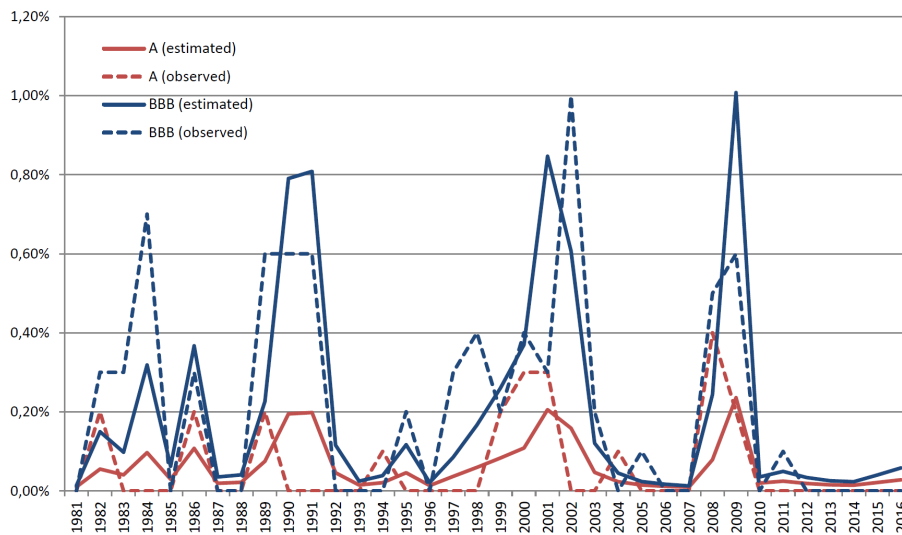


Figure 2: Estimated PIT PDs and observed default frequencies (1)

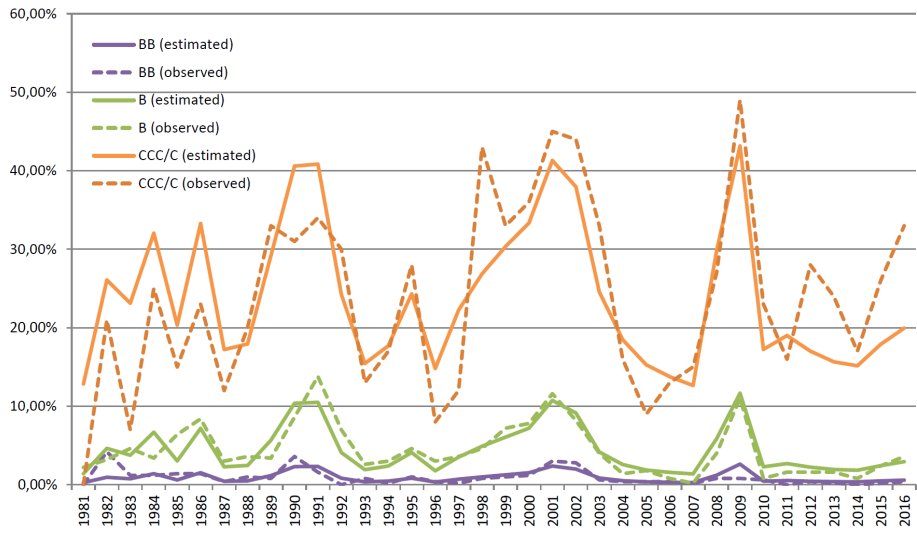


Figure 3: Estimated PIT PDs and observed default frequencies (2)

References

- Aguais, S., Forest Jr., L. R., King, L., Lennon, M. C., & Lordkipanidze, B. (2008). Designing and implementing a Basel II compliant PIT-TTC ratings framework. In *The Basel Handbook* (pp. 267–297). London: Risk Books, 2nd edition.
- Altman, E. I. & Rijken, H. A. (2004). How rating agencies achieve rating stability. *Journal of Banking and Finance*, 28(11), 2679–2714.
- Altman, E. I. & Rijken, H. A. (2006). A point-in-time perspective on through-the-cycle ratings. *Financial Analyst Journal*, 62(1), 54–70.
- Amato, J. D. & Furfine, C. H. (2004). Are credit ratings procyclical? *Journal of Banking and Finance*, 28(11), 2641–2677.
- Aussenegg, W., Resch, F., & Winkler, G. (2011). Pitfalls and remedies in testing the calibration quality of rating systems. *Journal of Banking and Finance*, 35, 698–708.
- Basel Committee on Banking Supervision (2005). Studies on the validation of internal rating systems. Bank for International Settlements, Working Paper 14.
- Basel Committee on Banking Supervision (2011). Basel III: A global regulatory framework for more resilient banks and banking systems. Bank for International Settlements.
- Basel Committee on Banking Supervision (2016). Consultative document: Reducing variation in credit risk-weighted assets - constraints on the use of internal model approaches. <http://www.bis.org/bcbs/publ/d362.htm>.
- Blümke, O. (2011). A proposal for a validation methodology for the discriminatory power of a rating system over time. *The Journal of Risk Model Validation*, 5(1), 21–44.
- Carey, M. & Hrycay, M. (2001). Parameterizing credit risk models with rating data. *Journal of Banking and Finance*, 25(1), 197–270.
- Carlehed, M. & Petrov, A. (2012). A methodology for point-in-time-through-the-cycle probability of default decomposition in risk classification systems. *The Journal of Risk Model Validation*, 6(3), 3–25.
- Cesaroni, T. (2015). Procyclicality of credit rating systems: How to manage it. *Journal of Economics and Business*, 82, 62–83.

- Coppens, F., Gonzáles, F., & Winkler, G. (2007). The performance of credit rating systems in the assessment of collateral used in eurosystem monetary policy operations. *ECB Occasional Paper Series*, No. 65.
- Coppens, F., Mayer, M., Millischer, L., Resch, F., Sauer, S., & Schulze, K. (2016). Advances in multivariate back-testing for credit risk underestimation. *ECB Working Paper Series*, 1885.
- Deutsche Bundesbank (2003). Approaches to the validation of internal rating systems. *Monthly Bulletin*, September, 59–71.
- European Banking Authority (2013). Summary report on the comparability and pro-cyclicality of capital requirements under the internal ratings based approach in accordance with article 502 of the capital requirements regulation. <https://www.eba.europa.eu/documents/10180/15947/20131217+Summary+report+on+comparability+and+pro-cyclicality+of+the+IRB+Approach.pdf>.
- European Banking Authority (2014). Guidelines on common procedures and methodologies for the supervisory review and evaluation process (SREP). <http://www.eba.europa.eu/documents/10180/935249/EBA-GL-2014-13+%28Guidelines+on+SREP+methodologies+and+processes%29.pdf>.
- European Banking Authority (2016). Consultation paper: Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures. <https://www.eba.europa.eu/regulation-and-policy/model-validation/guidelines-on-pd-lgd-estimation-and-treatment-of-defaulted-assets>.
- European Securities and Markets Authority (2015). Discussion paper: On the validation and review of credit rating agencies' methodologies. https://www.esma.europa.eu/sites/default/files/library/2015-1735_discussion_paper_on_validation_final.pdf.
- Hamilton, D. T., Sun, Z., & Ding, M. (2011). Through-the-cycle EDF credit measures. Moody's Analytics.
- Hodrick, R. & Prescott, E. (1981). Post-war u.s. business cycles: An empirical investigation. Discussion Paper 451, Northwestern University, Center for Mathematical Studies in Economics and Management Science.
- Hornik, K., Jankowitsch, R., Leitner, C., Lingo, M., Pichler, S., & Winkler, G. (2010). A latent variable approach to validate credit rating systems. In D. Rösch

- & H. Scheule (Eds.), *Model Risk in Financial Crises* (pp. 277–296). London: Risk Books.
- Hornik, K., Jankowitsch, R., Lingo, M., Pichler, S., & Winkler, G. (2007). Validation of credit rating systems using multi-rater information. *Journal of Credit Risk*, 3(4), 3–29.
- Ingolfsson, S. & Elvarsson, B. T. (2010). Cyclical adjustment of point in time PD. *Journal of the Operational Research Society*, 61(3), 374–380.
- Kiff, J., Kisser, M., & Schumacher, L. (2013). Rating through-the-cycle: what does the concept imply for rating stability and accuracy? *International Monetary Fund*, WP/13/64.
- Krahnert, J. P. & Weber, M. (2001). Generally accepted rating principles: A primer. *Journal of Banking and Finance*, 25(1), 3–23.
- Lingo, M. & Winkler, G. (2008). Discriminatory power - an obsolete validation criterion? *Journal of Risk Model Validation*, 4, 4–24.
- Loeffler, G. (2004). An anatomy of rating through the cycle. *Journal of Banking and Finance*, 28(3), 695–720.
- Loeffler, G. (2005). Avoiding the rating bounce: why rating agencies are slow to react to new information. *Journal of Economic Behavior and Organization*, 56(3), 365–381.
- Loeffler, G. (2013). Can rating agencies look through the cycle? *Review of Quantitative Finance and Accounting*, 40(4), 623–646.
- Mayer, M. & Sauer, S. (2017). Are through-the-cycle credit risk models a beneficial macro-prudential policy tool? In F. Heinemann, U. Klüh, & S. Watzka (Eds.), *Monetary Policy, Financial Crises, and the Macroeconomy*. Springer Verlag, Berlin.
- Merton, R. C. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29, 449–470.
- Petrov, A. & Rubtsov, M. (2016). A point-in-time-through-the-cycle approach to rating assignment and probability of default calibration. *The Journal of Risk Model Validation*, 10(2), 83–112.

- Prudential Regulation Authority (2013). Credit risk: internal ratings based approaches. Supervisory Statement SS1/13, <http://www.bankofengland.co.uk/prd/Documents/publications/ss/2013/ss113.pdf>.
- Prudential Regulation Authority (2017). Internal ratings based (IRB) approaches. Updated Supervisory Statement SS11/13, <http://www.bankofengland.co.uk/prd/Documents/publications/ss/2017/ss1113update.pdf>.
- Repullo, R., Saurina, J., & Trucharte, C. (2010). Mitigating the procyclicality of basel II. *Economic Policy*, 64, 659–702.
- Rösch, D. (2005). An empirical comparison of default risk forecasts from alternative credit rating philosophies. *International Journal of Forecasting*, 11, 37–51.
- Skoglund, J. (2017). Credit risk term-structures for lifetime impairment forecasting: A practical guide. *Journal of Risk Management in Financial Institutions*, 10(2), 177–195.
- Topp, R. & Perl, R. (2010). Through-the-cycle ratings versus point-in-time ratings and implications of the mapping between both rating types. *Financial Markets, Institutions & Instruments*, 19(1), 47–61.