Multi-State Markov Modeling of IFRS9 Default Probability Term Structure in OFSAA

ORACLE WHITE PAPER |

DARIO CZIRÁKY AND ULRICH ZINK

Disclaimer

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Table of Contents

Introduction

The expected credit loss (ECL) or impairment calculation rules imposed by the IFRS9 standard require the financial institutions to calculate expected loss for banking book facilities over the entire life of the exposures, conditional on macroeconomic factors, on a point-in-time basis. The new ECL calculation rules will apply from January 2018, by which time the banks will need to have them implemented in production IT systems. The IFRS9 methodology for ECL calculation requires models for probability of default (PD), loss given default (LGD) and exposure at default (EAD). In this paper we describe a PD term structure model based on multi-state Markov (MSM) methodology. The inputs to the model are discrete rating grades that come from either bank's internal rating system or from the rating agencies, and macroeconomic time series. The model produces cumulative PDs over multiple tenor points, conditional on specific values of the macroeconomic factors (macro scenarios). We then show how the estimated model can be implemented in Oracle's OFSAA, namely its IFRS9 module. We estimate and deploy the models in Oracle using the R programming language.

IFRS9 standards for PD modeling

The expected credit loss (ECL) or impairment calculation rules set by the IFRS9 standard require the financial institutions to calculate expected loss for banking book facilities over the entire life of the exposures, conditional on macroeconomic factors on a point-in-time basis. The new ECL calculation rules will apply from January 2018, by which time the banks will need to have the ECL calculations implemented in production IT systems.

In this section we briefly outline the main requirements for the estimation of default probabilities from the IFRS9 Standard point of view.

- PD estimates should be unbiased ("best estimate PD"), i.e. PD should accurately predict number of defaults and does not include optimism or conservatism. Where regulatory capital models are used as a starting point to calculate ECL, appropriate adjustments need to be made to remove inherent conservatism.
- Estimated PDs should be point-in-time, i.e., adjusted, where necessary, to reflect the effects of the current economic conditions.
- The PD estimates should be recalibrated on a regular basis, or monitoring should be provided to show why recalibration was not necessary. The (re)calibration sample should representative of the current population.
- PD should be calculated using sufficient sample size, i.e. estimates should be based on a sufficiently large sample size to allow for a meaningful quantification and validation of the loss characteristics. Historical loss data should cover at least one full credit cycle.
- The PDs should be calculated with appropriate segmentation - the bank should consider risk drivers in respect of borrower risk, transaction risk and delinquency status in assigning exposures to PD model segments.
- The data used for calibration should be consistent with the IFRS 9 default definition, i.e. all components of the PD calculation (including 12m PDs and default curves for extrapolation to lifetime) should use the same definition of default.
- Lifetime PDs should be calculated using appropriate extrapolation methodologies. Where extrapolation techniques are used to determine lifetime PD measures, these should not introduce bias into the calculation of ECL.
- The estimated PDs should be inclusive of forward looking information, including macroeconomic factors in the computation of lifetime PD's, to ensure that loss recognition is not delayed.

- For instruments that have comparable credit risk, the risk of a default must be higher the longer the expected life of the instrument (this requires that cumulative lifetime PD curves are monotonically increasing)
- Internal data should be used in building the PD models, where possible, and data should be representative of the portfolio going forward.
- Where external data or vendor models are used, the external calibration sample should be representative of the internal risk profile of the current population.

The practical implications of these requirements is that PD term structure models need to be quite sophisticated and should fit the data very well to provide credibility in estimating long tenor PDs over the life of the exposures that can often be in excess of 20 years.

The AIRB requirements for PD models used for capital calculations differ from IFRS9 in two main respects. First, The AIRB PDs can be TTC, PIT, or hybrid with an additional requirement (in CRD4) that capital should not be pro-cyclical, hence acceptable PD models and rating systems that produce PD parameters for capital calculations tend to be TTC or hybrid. The exception is PIT PD models that have an additional adjustment mechanism (e.g. variable scalar approach) that converts PIT to TTC PDs for capital calculation purposes. In such case the Basel and IFRS9 1-year PDs will be equivalent and there is no need for further adjustment, however this case is not common in practice and there is only one major bank using PIT PD estimation for sanctioning purposes and an auxiliary adjustment framework that results in TTC PDs for capital calculations. Most banks will need to adjust their rating system from a TTC/hybrid to PIT to make their internal (1-year) PDs compatible with IFRS9.

The second difference between AIRB/Basel and IFRS9 requirements is the IFRS9 requirement for PDs to be unbiased, or “best estimate”, which contrasts to the Basel requirement for PDs to be “conservative”. This poses challenges for the development of IFRS9 impairment models. Namely, the development of AIRB/IRB PD models typically made use of the conservative regulatory guidance to cover for any data or model deficiency by simply increasing PD estimates, adding floors or modifying any other model parameters that will have the end results of increasing risk weighted assets and hence capital. This means that many AIRB models already approved by the regulators will not be suitable for IFRS9 without further modifications. Specifically, any conservative overlay needs to be removed in order to use the PDs in impairment calculation under IFRS9. This can be tricky as conservativism is likely to have been introduced to remedy model deficiencies, e.g. poor rank ordering or imprecisely estimated model coefficients, hence once removed these deficiencies will become a model weakness thereby calling for a complete model rebuild.

Overall, the biggest challenge for PD modeling coming from IFRS9 is to link the defaults (or more broadly rating migrations) to macroeconomic variables and to do so over a very long time horizon. The AIRB/Basel PD models aimed to forecast defaults over one year horizon, which was typically done on the long-term average, through-the-cycle, basis. Performance of TTC PD models is assessed over a long time period, “on average”; hence under/over-prediction in particular years is not a sign of model under-performance. On the other hand, the PD forecasts for IFRS9 impairment calculation are point-in-time, conditional on macroeconomic variables, and are forecasted over multiple years. Such models thus need to have good empirical performance in each year, at different points in the cycle, and they should be accurate rather than conservative.

In this paper we estimate a panel multi-state Markov model using discrete credit ratings data combined with macroeconomic time series. Section 2 describes the data set and section 3 outlines how TTC ratings are adjusted to PIT ratings before model is estimated. Section 4 describes the MSM methodology while model estimates are reported in section 5. In section 6 we show how conditional lifetime PD curves can be calculated from an MSM model. Section 7 shows how the model estimated in R language can be deployed to production from Oracle’s OFSAA IFRS9 core banking software solution.

Data

The data for modeling the term structure of conditional default probabilities consists of obligor rating histories and macroeconomic time series. An appropriate data format is a panel (cross-sectional time series) where the cross section is across obligors and time series is over time. It would be assembled with obligor ratings (internal or form a rating agency) for each date (time point) when the rating was changed along with the associated value of macroeconomic variables at the same time point. For example, suppose an obligor rating was downgraded from BBB to BB- in December 2009, at the same time the change in employment in the United Kingdom compared to one year earlier was approximately -2%, hence this information forms one single observation point in the panel data set. Other macroeconomic variables would enter in the same way, i.e. annual change as of December 2009 (or level if a variable is not differenced).

For the purpose of illustrating the MSM modeling for IFRS9 we use UK macroeconomic time series spanning from 2007 to 2014, which resembles data span most British banks, would be using for such purposes. We consider GDP, employment, consumption, and FTSR100 (see Figure 1). Except for employment, all variables are in annual differences (changes) as their levels are non-stationary.

In practice, the choice of macro factors will depend on the nature of the portfolio and different factors would generally be needed for different types of portfolios and industry sectors. The ratings data source can be external (Moody's, S&P, Fitch) or bank's internal ratings history. For the purpose of illustrating the modeling process we make use of simulated ratings history data on a simplified rating scale with 5 performing grades (G1-G5) and a default grade (G6).

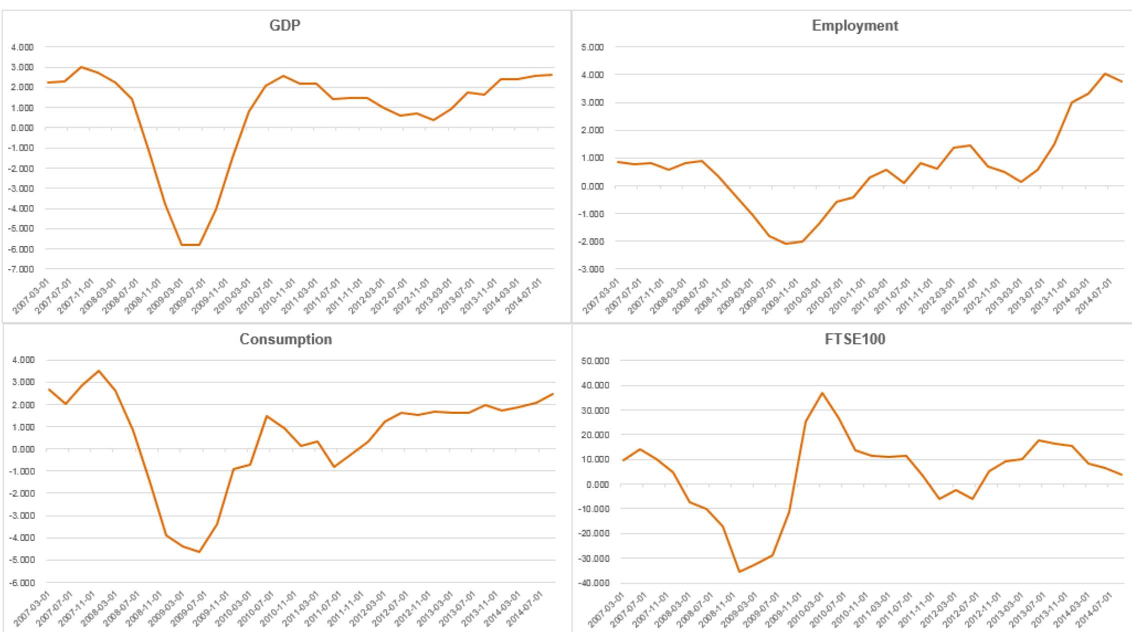

Figure 1. Macroeconomic time series – monthly changes

Point-in-Time Adjustment of Through-the-Cycle Ratings

Consider an example of S&P ratings, which are largely through-the-cycle (TTC). S&P reports an empirical default rate for BBB grade calculated as an annual average from 1981 to 2014 as 0.32. This is clearly a TTC default probability as it was calculated over a 33 year period. In each of these years the observed BBB default rate was generally either below or above the long run of 0.32. This annual default rate calculated for each individual year is the point-in-time (PIT) default rate of the BBB grade. This means that implied PIT rating for a company that has TTC BBB long-term rating can be e.g. 0.12 in any given year. The 0.12 default rate in the long-run corresponds to A grade hence in a good year a long-run BBB can be represented by an A grade if its observed default rate is, say, 0.12 in that year. In this case, the TTC rating would be BBB and PIT rating would be A. Such conversion mapping can be used across all grades and years to produce PIT ratings.

Mathematically, the cohort estimate of the TTC default rate for a generic rating grade j over all time periods t can be expressed as a simple ratio

$$P\left(G_j^{TTC} = D\right) \equiv \frac{D_{G_j \to D}}{N_{G_j}}$$

while in a specific time period (e.g. year) from time t_1 to time t_2, the default rate is calculated as

$$P\left(G_j^{PIT} = D \mid t_1 < t < t_2\right) \equiv \frac{\underset{t_1<t<t_2}{D_{G_j \to D}}}{\underset{t_1<t<t_2}{N_{G_j}}}$$

Using our sample data set we calculate TTC transition matrix over the 2007-14 period based on equation (1), which gives results shown in Table 1.

TABLE 1. THROUGH-THE-CYCLE TRANSITION MATRIX

	G1	G2	G3	G4	G5	D
G1	0.812	0.162	0.018	0.001	0.004	0.004
G2	0.158	0.761	0.058	0.011	0.004	0.008
G3	0.020	0.234	0.542	0.103	0.050	0.050
G4	0.010	0.160	0.144	0.433	0.144	0.108
G5	0.007	0.109	0.036	0.109	0.467	0.270

Applying equation (2), i.e. calculating specific year default rate for each rating grade and then mapping the observed default frequency to TTC grades for 2008 gives the results shown in Table 2. The resulting grades G1-G5 are now PIT.

TABLE 2. 2008 POINT-IN-TIME TRANSITION MATRIX

	G1	G2	G3	G4	G5	D
G1	0.807	0.114	0.068	0.000	0.011	0.000
G2	0.082	0.776	0.102	0.031	0.000	0.010
G3	0.023	0.205	0.545	0.114	0.068	0.045
G4	0.077	0.077	0.000	0.462	0.000	0.385
G5	0.000	0.000	0.000	0.000	0.500	0.500

For comparison, we do the same calculation for year 2014, to give us a balance of bad and good years, which is shown in Table 3.

TABLE 3. 2014 POINT-IN-TIME TRANSITION MATRIX

	G1	G2	G3	G4	G5	D
G1	0.851	0.142	0.000	0.000	0.000	0.007
G2	0.085	0.858	0.019	0.028	0.009	0.000
G3	0.000	0.188	0.688	0.000	0.063	0.063
G4	0.000	0.333	0.167	0.333	0.000	0.167
G5	0.000	0.200	0.000	0.000	0.400	0.400

We can observe that the dynamics in the transition matrices above are quite different specially the default rates. Note the default rate associated with G4 in 2008 is 0.385, which corresponds to G5 in the long-run; hence a G5 obligor through-the-cycle is actually a G5 point-in-time in 2008. This means that PIT credit quality (grade) in 2008 is lower than the TTC grade, for the same obligor. The opposite holds in good years, as in our 2014 example in Table 3 above shows a G3 obligor in 2014 has credit quality that is better than the long-run (TTC) default rate for G3, and hence maps into G2 or G1 (is nearly half way in between in this case). We thus follow this logic and map yearly default rates into the TTC-defined grades, which gives us PIT grades as illustrated in the above example. The PIT grades data will be used in our modeling as this will give us correct transition matrix dynamics and correlations with the macro factors.

Methodology

We use a multi-state Markov model (MSM) to model transition probabilities between rating grades (see Cziráky and Sadleir, 2013). A multi-state Markov model captures movements between n states where the probability of moving away from the current state depends on the previous state. We consider transitions between states in continuous time, measured by an $\boldsymbol{n \times n}$ intensity matrix $\boldsymbol{Q(t)}$ given by

$$\mathbf{Q}(t,\mathbf{z}(t)) = \begin{pmatrix} -\sum_{j\neq 1}^{n-1} q_{1j}(t,z(t)) & q_{12}(t,z(t)) & \cdots & \cdots & q_{1n}(t,z(t)) \\ q_{12}(t,z(t)) & -\sum_{j\neq 2}^{n-1} q_{2j}(t,z(t)) & \cdots & \cdots & q_{2n}(t,z(t)) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ q_{n-1,1}(t,z(t)) & q_{n-1,2}(t,z(t)) & \cdots & -q_{n-1,n-1}(t,z(t)) & q_{n-1,n}(t,z(t)) \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \quad (1)$$

where the instantaneous risk of going from state *r* to state *s* is represented by the intensity

$$q_{rs}(t,z(t)) = \lim_{\delta t\to 0} \frac{P(S(t+\delta t)=s \mid S(t)=r)}{\delta t}, \quad (2)$$

which can be written as $q_{rs}(t,z(t)) = q_{rs}$ for time-homogeneous models,

$$\ln(q_{rs}(z_{ij})) = \ln(q_{rs}^{(0)}) + \sum_{k=1}^{K} \beta_{rs,k} x_k , \quad (3)$$

or equivalently

$$q_{rs}(z_{ij}) = q_{rs}^{(0)} \exp\left(\sum_{k=1}^{K} \beta_{rs,k} x_k\right), \quad (4)$$

Where x_k, $k = 1, 2, \ldots, k$ denotes a set of exogenous variables affecting the instantaneous risk of going from state r to state s. In this case we have the same set of K covariates affecting different transitions, while the intercept ***ln** (q_{rs})* is specific to the transition intensity from state s to state r.

When panel data are considered we have *n x n* states, *m* obligors observed over *T* time periods (see Kalbfleisch and Lawless (1985) for more details on Markov models for panel data). If panel is balanced, each of the *m* obligors will be observed in all *T* time periods, otherwise, in the case of an unbalanced panel design some of the *m* obligors will be observed for a period shorted than *T*. Starting times might generally differ as well.

The likelihood of the MSM model for panel data can be written in terms of transition probabilities p_{rs} (denoting probability of transition from state r to state s) by multiplying across all states, obligors, and time periods

$$L(\mathbf{Q}(t,\mathbf{z}(t))) = \prod_{r=1}^{n}\prod_{s=1}^{n}\prod_{i=1}^{m}\prod_{t=1}^{T} p_{r,s,i,t} \quad (5)$$

The likelihood is parameterized in terms of transition probabilities, which form an ***n x n*** transition probability matrix ***P (t,z(t))*** corresponding to the intensity matrix ***Q(t,z(t))*** that can be written as

$$\mathbf{P}(t,\mathbf{z}(t)) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}, \quad (6)$$

where the relationship between ***P(t,z(t))*** corresponding to the intensity matrix ***Q(t,z(t))*** matrices is given by the following equation

$$\mathbf{P}(t,\mathbf{z}(t)) = \exp\left(t\mathbf{Q}(t,\mathbf{z}(t))\right) = \sum_{j=0}^{\infty} \frac{t^j \mathbf{Q}(t,\mathbf{z}(t))^j}{j!} \quad (7)$$

The estimation of the model coefficients is done by maximum likelihood, which is available in the analytical form for simple models with small number of states (Kalbfleisch and Lawless, 1985). In general, for models with larger number of states ($n > 3$), optimization methods using numerical derivatives and derivatives-free techniques need to be used (see Jackson, 2011).

Model Specification and Estimation

We model the transition probabilities between pairs of states represented by 5 letter grades (G1-G5), where G6 denotes default. For the estimation purposes we allow all possible transitions, i.e. we do not impose any zero constraints on the ***Q(t,z(t))*** matrix.

We model the dependence of the transition probabilities (or equivalently intensities) to macroeconomic variables shown in Table 4.

TABLE 4. VARIABLES AND NOTATION

Macroeconomic variable	Notation
GDP (annual change)	ΔGDP_t
Employment rate	emp_t
Consumption (annual change)	$\Delta cons_t$
FTSE100 (annual change)	ΔS_t

In the estimation we consider four model specifications where the link function (4) will be parameterized as

$$\text{Model 1: } \sum_{k=1}^{K} \beta_k x_k = \beta_1 \Delta gdp_t, \quad K=1$$

$$\text{Model 2: } \sum_{k=1}^{K} \beta_k x_k = \beta_1 \Delta gdp_t + \beta_2 emp_t, \quad K=2$$

$$\text{Model 3: } \sum_{k=1}^{K} \beta_k x_k = \beta_1 \Delta gdp_t + \beta_2 emp_t + \beta_3 \Delta cons_t, \quad K=3$$

$$\text{Model 4: } \sum_{k=1}^{K} \beta_k x_k = \beta_1 \Delta gdp_t + \beta_2 emp_t + \beta_3 \Delta cons_t + \beta_4 \Delta s_t, \quad K=4$$

We estimate the model parameters by maximum likelihood using BFGS quasi-Newton algorithm. In particular, we estimate nested models with an increasing number of exogenous variables (Models 1-4) and a base-case model without any exogenous variables (Model 0). In each of the estimated models we allow transition probabilities to depend on all included macroeconomic variables, which can be relaxed by imposing parameter constraints in the estimated model.

Significance of sequential inclusion of individual exogenous variables is tested by the likelihood-ratio test with the results reported in Table 5. We find that the inclusion of ΔGDP_t significantly improves the model over the base case with no covariates. Further inclusion of emp_t and $\Delta cons_t$ did not lead to improvement while adding ΔS_t leads to significant improvement of the previous models.

TABLE 5. MODEL SPECIFICATION TESTING

Model testing	LR value	d.f.	p-value
Model 1 vs. Model 0	63.25801	25	0.000
Model 2 vs. Model 1	26.27585	25	0.393
Model 3 vs. Model 2	29.0173	25	0.263
Model 4 vs. Model 3	61.91031	25	0.000
Model 4 vs. Model 1	117.2035	75	0.001

Table 6 reports full set of parameter estimates along with their 95% confidence bounds.

To evaluate the model performance we compare the Model 4's predicted survival curves to a nonparametric (Kaplan-Meier) curve, which is shown in Figure 2. We find that the model fits well between 1 and 3 years and it then over-predicts the empirical survival rate beyond 3rd year. Given the short time series for most obligors in the panel data set we cannot expect good forecasting performance at the longer-horizons due to scarcity of data, which holds for both MSM model and empirical Kaplan-Meier estimates. However, for typical wholesale products[1] the maturity is mostly under 5 years hence this need not be a model weakness.

[1] E.g. revolving loans, bullet loans, letters of credit, etc.

In addition to comparing empirical to model-predicted survival rates, we can also compare predicted and observed prevalence of states over time. Prevalence measures the proportion of obligors rated by each of the 5 rating grades among obligors in all grades, hence such comparison looks at how well the model predicts the proportion of obligors in each rating bucket, which is shown in Figure 3.

TABLE 6. MLE PARAMETER ESTIMATES

Transition	Base	Lower	Upper	GDP	Lower	Upper	Employment	Lower	Upper	Consumption	Lower	Upper	FTSE100	Lower	Upper
1 - 1	-0.238	-0.265	-0.213	--	--	--	--	--	--	--	--	--	--	--	--
1 - 2	0.220	0.196	0.246	0.826	0.730	0.936	1.121	0.940	1.336	1.078	0.942	1.233	1.013	1.002	1.025
1 - 3	0.011	0.006	0.020	1.092	0.647	1.844	0.344	0.155	0.763	1.648	0.967	2.807	0.912	0.868	0.959
1 - 4	0.000	0.000	19.378	0.106	0.001	17.574	0.308	0.003	34.342	83.307	0.032	>100.0	0.655	0.293	1.466
1 - 5	0.003	0.001	0.009	1.248	0.459	3.397	0.695	0.141	3.430	1.101	0.400	3.029	0.931	0.844	1.026
1 - 6	0.004	0.002	0.010	0.828	0.332	2.069	0.628	0.153	2.583	1.609	0.585	4.424	0.974	0.899	1.055
2 - 1	0.222	0.197	0.249	1.013	0.895	1.147	1.216	1.024	1.443	0.929	0.813	1.061	1.007	0.996	1.018
2 - 2	-0.321	-0.354	-0.290	--	--	--	--	--	--	--	--	--	--	--	--
2 - 3	0.072	0.058	0.090	1.012	0.805	1.272	0.651	0.445	0.952	1.172	0.917	1.498	0.974	0.953	0.995
2 - 4	0.011	0.006	0.020	0.744	0.439	1.262	0.579	0.258	1.300	2.197	1.199	4.025	0.981	0.939	1.024
2 - 5	0.005	0.002	0.012	1.036	0.432	2.487	0.555	0.123	2.511	1.050	0.414	2.663	0.985	0.905	1.071
2 - 6	0.011	0.006	0.018	0.982	0.547	1.761	1.090	0.444	2.678	0.902	0.476	1.711	0.996	0.942	1.054
3 - 1	0.023	0.009	0.054	1.162	0.485	2.785	0.895	0.232	3.459	0.786	0.310	1.995	0.977	0.900	1.060
3 - 2	0.405	0.339	0.484	0.994	0.817	1.209	1.425	1.141	1.779	0.869	0.700	1.080	1.017	1.000	1.033
3 - 3	-0.710	-0.818	-0.617	--	--	--	--	--	--	--	--	--	--	--	--
3 - 4	0.131	0.092	0.188	1.066	0.732	1.555	0.658	0.361	1.201	0.942	0.646	1.375	0.998	0.965	1.031
3 - 5	0.070	0.044	0.112	1.232	0.739	2.053	0.836	0.389	1.798	0.840	0.501	1.409	0.980	0.937	1.025
3 - 6	0.081	0.054	0.122	0.856	0.541	1.354	1.130	0.627	2.038	1.133	0.690	1.861	0.998	0.960	1.036
4 - 1	0.004	0.000	0.228	0.719	0.063	8.157	1.851	0.258	13.293	3.713	0.365	37.816	0.921	0.746	1.137
4 - 2	0.317	0.219	0.460	0.818	0.551	1.215	1.557	0.950	2.554	1.034	0.673	1.589	1.018	0.984	1.054
4 - 3	0.240	0.150	0.385	1.225	0.754	1.990	0.884	0.426	1.834	0.917	0.578	1.457	0.987	0.946	1.030
4 - 4	-0.933	-1.179	-0.739	--	--	--	--	--	--	--	--	--	--	--	--
4 - 5	0.251	0.162	0.389	0.778	0.514	1.176	2.076	1.268	3.398	0.793	0.487	1.291	1.048	1.011	1.086
4 - 6	0.121	0.057	0.257	1.568	0.773	3.182	0.334	0.114	0.981	1.316	0.749	2.312	0.914	0.857	0.974
5 - 1	0.000	0.000	>100.0	0.110	0.000	>100.0	0.557	0.000	>100.0	5.038	0.000	>100.0	1.167	0.248	5.477
5 - 2	0.198	0.114	0.347	1.028	0.597	1.771	1.066	0.449	2.531	0.917	0.506	1.661	1.001	0.945	1.061
5 - 3	0.025	0.003	0.212	0.559	0.192	1.631	0.802	0.169	3.810	3.227	0.851	12.242	1.054	0.930	1.195
5 - 4	0.154	0.075	0.316	1.001	0.558	1.795	1.444	0.541	3.853	0.606	0.283	1.300	1.021	0.955	1.092
5 - 5	-0.877	-1.153	-0.667	--	--	--	--	--	--	--	--	--	--	--	--
5 - 6	0.499	0.352	0.707	1.052	0.752	1.473	1.198	0.730	1.965	0.893	0.616	1.295	0.999	0.964	1.036

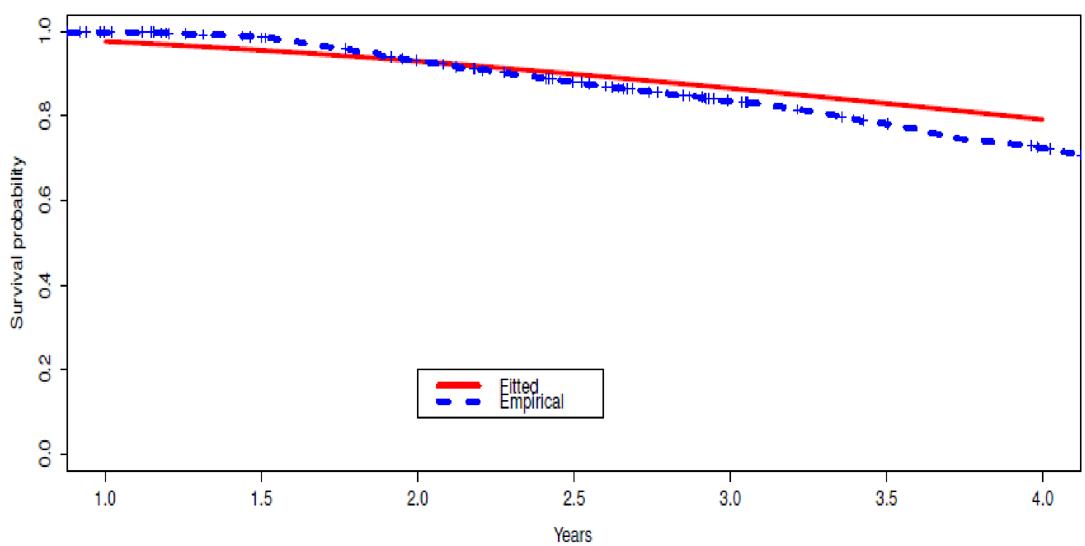

Figure 2. Survival from 1 to 4 years

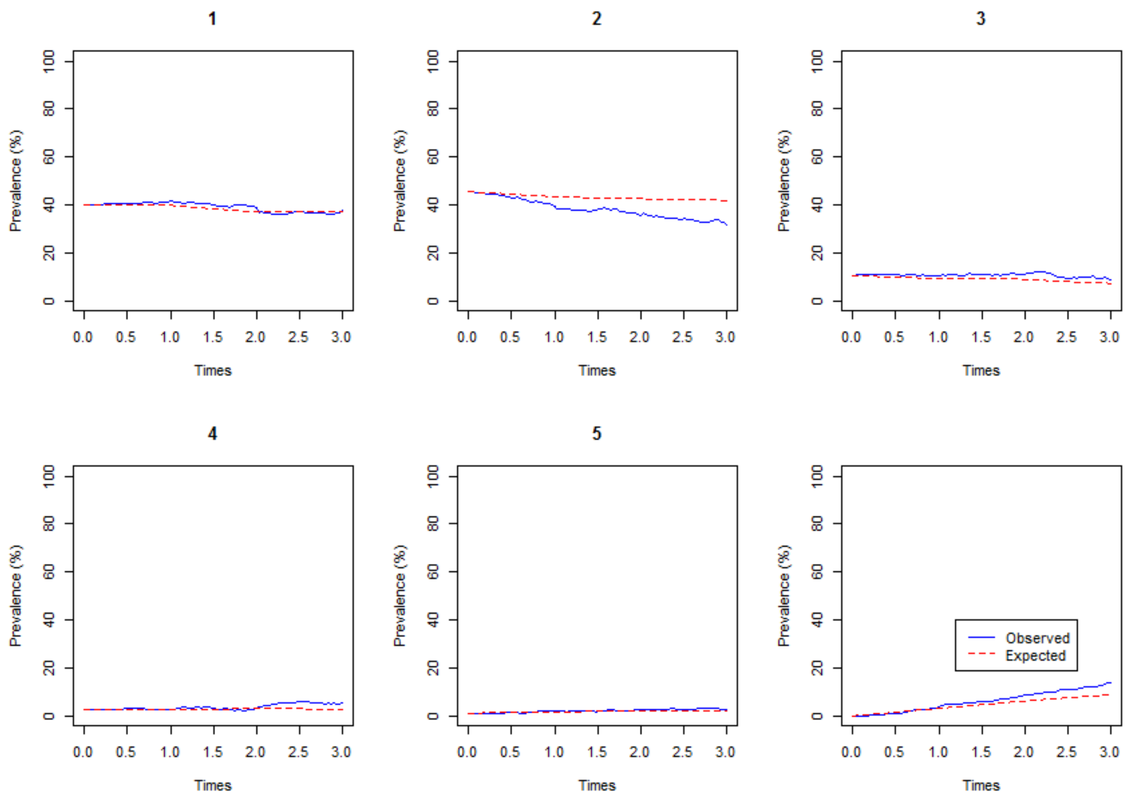

Figure 3. Prevalence years 1 to 3: Model 4

Calculation of Lifetime PDs Conditional on Macro Factors

The IFRS9 calculation of expected credit loss (ECL) requires marginal PD curves over the life of the exposures, conditional on the macroeconomic factors. In this section we outline calculation of cumulative conditional PD curves. We note that ECL calculation uses marginal PD curves, which are just one-period (e.g. annual) changes in the cumulative PD curves.

The one year-ahead cumulative PDs are in the last column of the conditional transition matrix that is calculated by entering specific values of the macroeconomic factors (scenarios or forecast) in the transition intensity formula for each transition state, i.e. each element of the intensity matrix. The matrix is then exponentiated which yields the transition probability matrix. The second year cumulative conditional PDs are obtained by doing the same thing with two years-ahead macro forecasts and then pre-multiplying the resulting transition matrix with the year-one matrix obtained in the previous step. This then repeats to *h* years, which can be expressed in general by the following formula

$$\mathbf{CP}\left(t,\mathbf{z}(t)\right)=\prod_{j=1}^{h}\mathbf{P}\left(t,\mathbf{s}\left(t=j\right)\right) \tag{8}$$

To demonstrate how the calculation works, we produce a hypothetical macro scenario for each of the four macro factors in the model over the next 21 years, as shown in Figure 4.

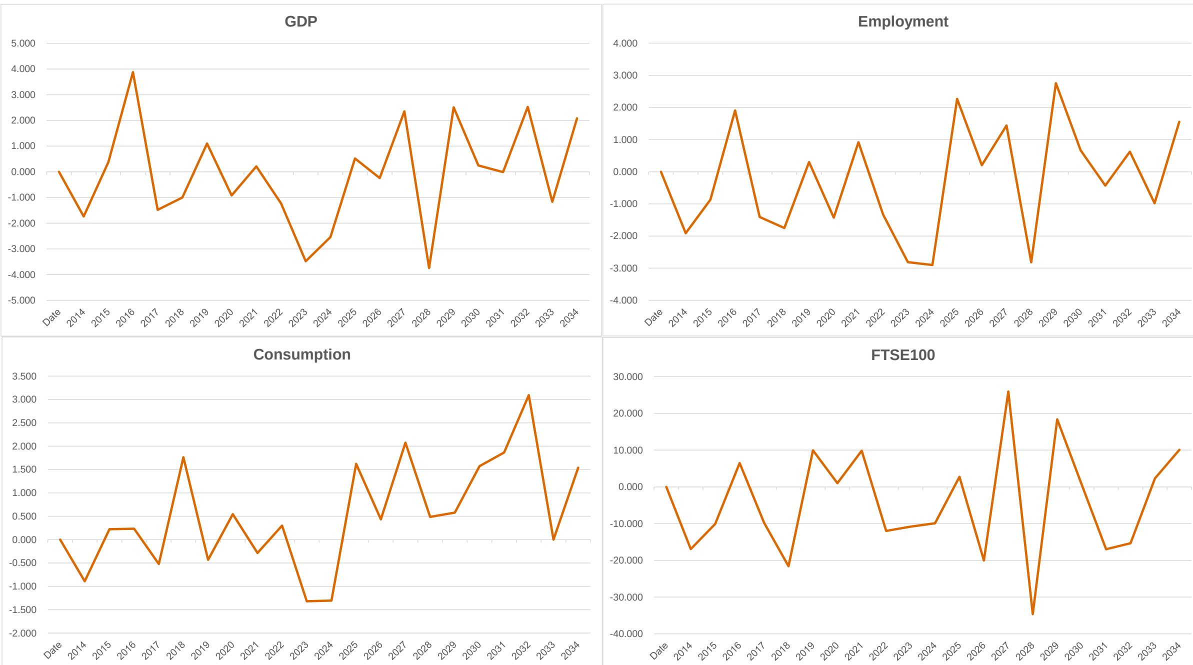

Figure 4. Hypothetical 21 year scenario

This hypothetical example results in cumulative PD curves shown in Figure 5. We note that default probabilities for the middle grades (G3 and G4) have notably increased, which is in line with a negative outlook for most of the variables in scenario above.

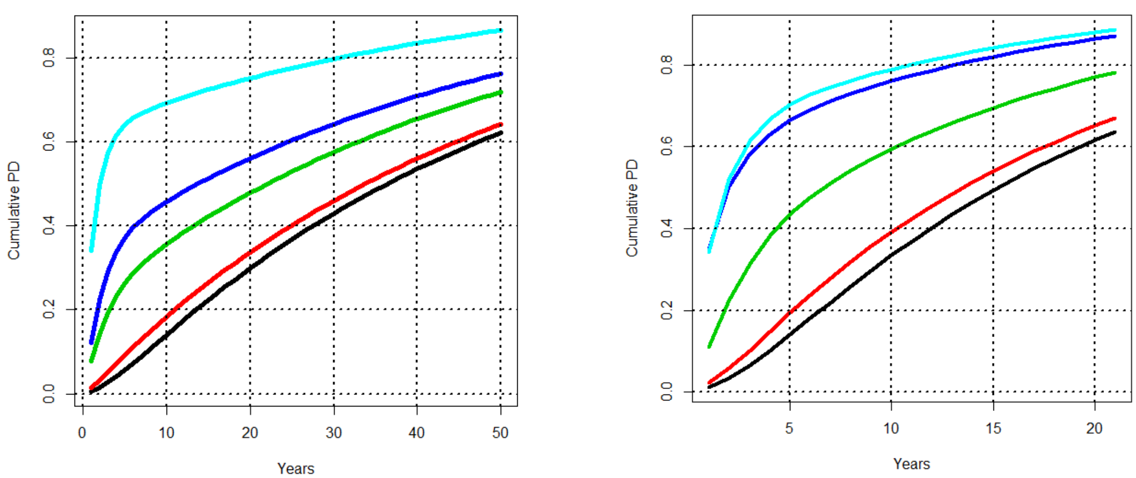

Figure 5. Base case (left) and conditional (right) cumulative PD curves (from 1 to 50 years)

At this point, what is important to note is that PD curves need to be responsive to macroeconomic scenarios as this is the ultimate aim of the PD term structure model, hence it should be carefully assessed how the conditional curves behave under different macroeconomic scenarios.

Production System Implementation of MSM PD Model in Oracle

In this section we describe how the estimated MSM PD model can be implemented in the Oracle Financial Services Analytical Applications' (OFSAA) IFRS9 solution. OFSAA is a unified suite of products that offer a framework for integrated risk and performance management. The OFSAA framework shares a common financial services data model, analytical infrastructure and business intelligence layer as depicted in Figure 6. This infrastructure delivers unified metadata across the stack and provides a single set of computational engines, stochastic modeling methods and business rules to feed overlapping, but independent, analytical business functions.

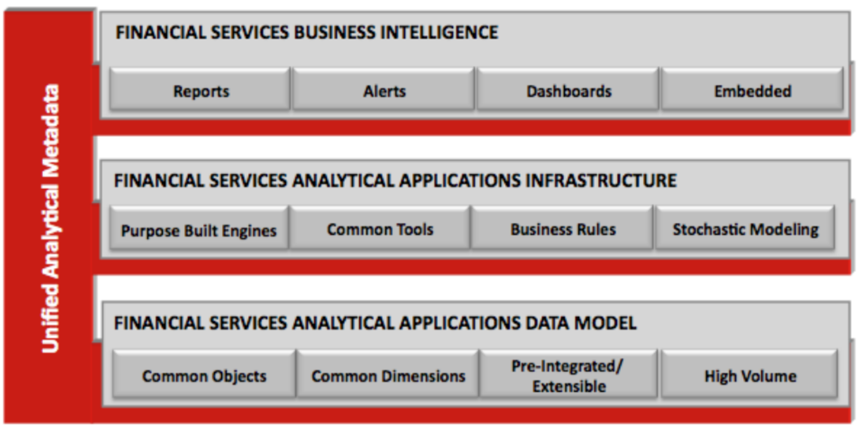

Figure 6. Building Blocks: Shared Components across Applications

These common "building blocks" are leveraged from application to application to ensure data consistency, traceability and availability throughout the enterprise.

As part of the OFSAA framework, the OFSAA IFRS9 solution includes the core IFRS9 functionalities for stage allocation and ECL calculation. It also manages the IFRS9 process workflow from data sourcing through to reporting, and comes with pre-configured methods for estimating loan loss provisions that enable compliance with IFRS 9 regulations, covering:

(i) Data management
(ii) Amortized cost and Effective Interest Rate (EIR) calculations
(iii) Impairment allowance & provisioning
(iv) Accounting generation
(v) Reporting

In addition, in order to provide full customization and more advanced modeling capabilities, the solution includes the Oracle Financial Services Enterprise Modeling module which allows statistical models written in various programming languages to be deployed straight into the OFSAA production environment.

OFSAA Enterprise Modeling suite comes with a selection of prebuilt ECL models; however, it also allows fast implementation of internally developed specific models regardless of their complexity. Oracle provides a toolkit for developing end-to-end analytical applications with data lineage and traceability enabled at every step along the analytical workflow, supporting techniques developed using R, C++ or Java or Python programming languages.

In this paper we make use of the R programming language[2] to develop and deploy the model within Oracle R Enterprise (ORE), an in-database implementation of the R platform. ORE is a component of the Oracle Advanced Analytics Option of Oracle Database Enterprise Edition. ORE enables running R models within the Oracle Database Environment. OFSAA supports ORE based models, that is, models defined and registered within the platform may be executed on an ORE instance. R models can be coded up within the platform or may be imported into the platform from the existing R scripts.

The deployment of R script that was either developed externally (e.g. in R Studio code editor) or within OFSAA application is carried out in three simple steps:

(i) Importing the script from the prototyping environment (R-studio) into OFSAA
(ii) Mapping the input variable with the corresponding elements of the data model in the database
(iii) Mapping the output variable with the corresponding elements of the data model in the database

The R code shown in OFSAA model script window in Figures 7 is created using the MSM R package (Jackson, 2011) and prototyped in R Studio running R engine version 3.3.2 on 64-bit. To run the code in OFSAA using ORE Engine we write a declaration that defines variables, model input and output. Note the "##" is not commenting the code out within the declaration section, it is a command that tells the underlying machine code to parse the R commands that follow. In the reaming of the text the "#" symbol serves the usual purpose of a comment marker.

At this point we do not go into details on how the R code is written, we just note once the declaration is written linking inputs and outputs to named R objects the rest of the code is exactly the same as the code we would write for desktop execution in standard R language.

The purpose of linking R objects with generic input/output declarations is to establish the link with the data base. Once the objects are declared, the model script will read the relevant fields from the data base (Figure 8), which requires filling in the model mapping dialog. Finally, as shown in Figure 9, the final dialog allows the user to store new tables in the data base containing model outputs such as the calculated ECL numbers, for each exposure, or intermediate model output as needed for audit or data retention purposes.

[2] R is an open source statistical programming language and environment for computing and graphics. For more information about R, see the R Project for Statistical Computing at http://www.r-project.org.

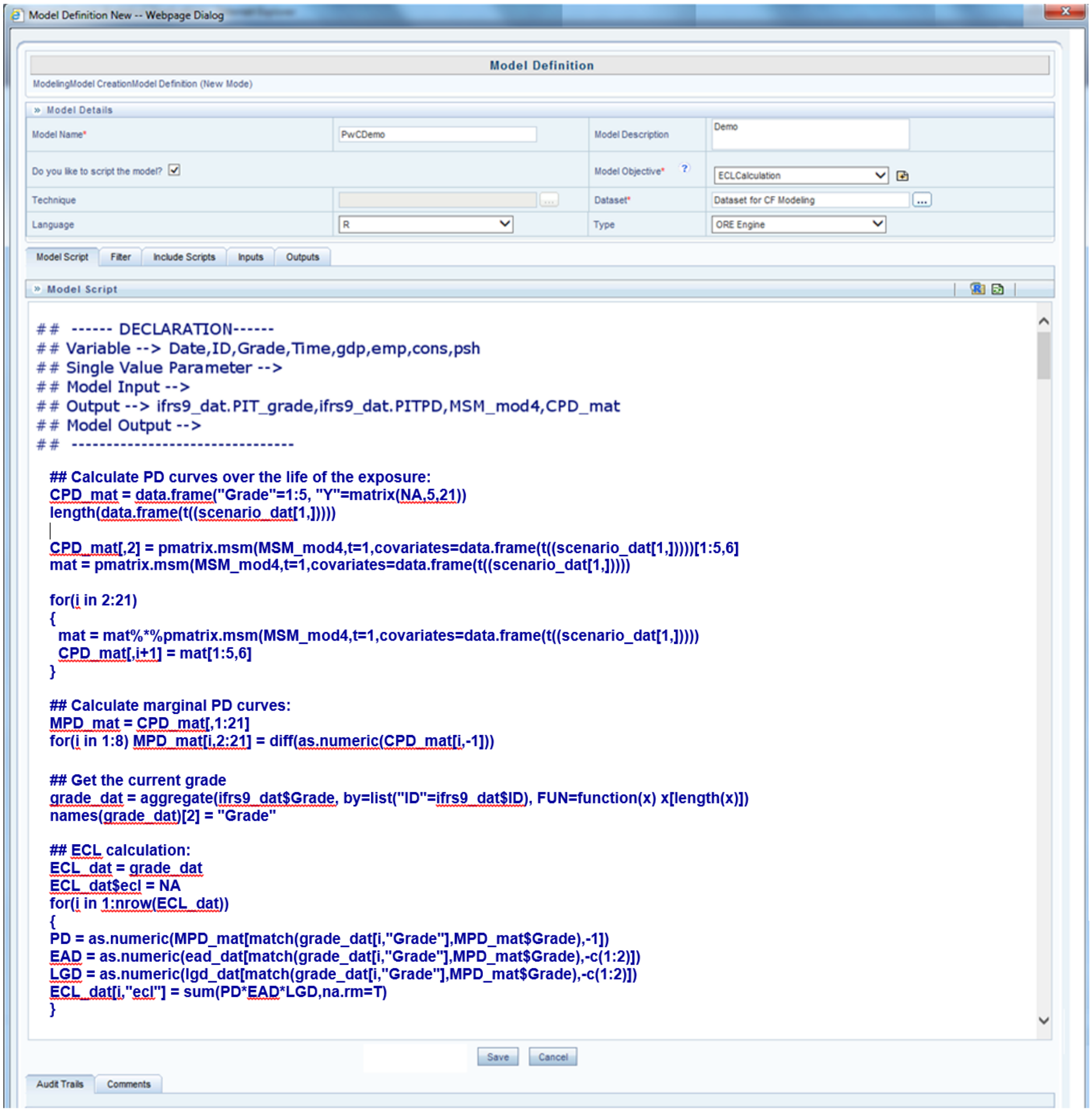

Figure 7. Model Input Screen

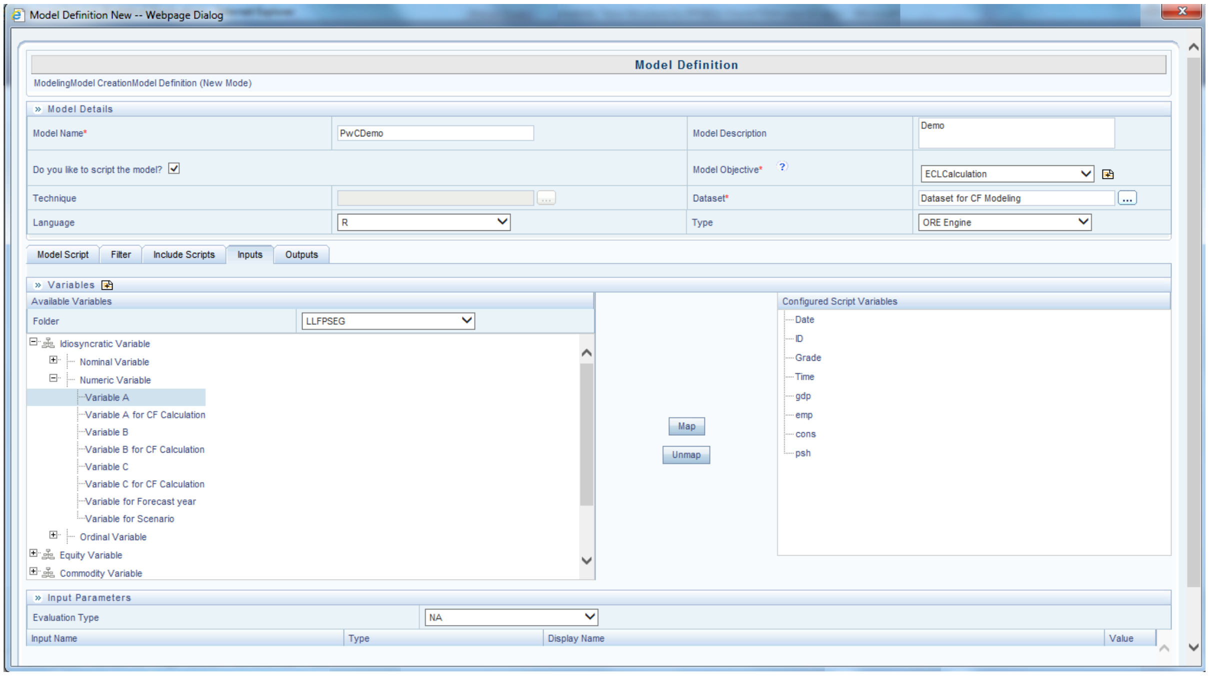

Figure 8. Input variable Mapping

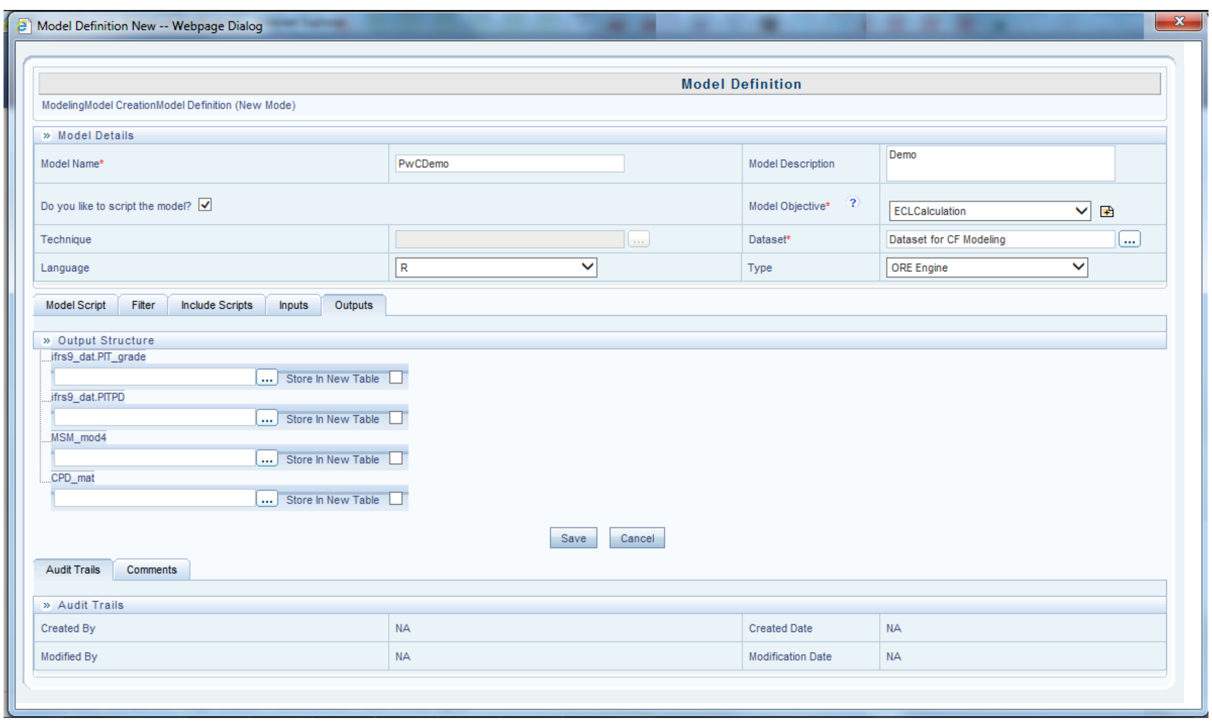

Figure 9. Output Variable Mapping

Conclusion

In this paper we show how a multi-state Markov model can be estimated and used for IFRS9 impartment purposes. This approach allows conditioning on macroeconomic factors, within the same model along with testing for model fit and significance of the estimated coefficients. We then show how the model estimated in R language can be implemented in Oracle's IFRS9 solution, which is part of the OFSAA suite of software. We show that once the R code is written, very little further effort is needed to deploy it in the production (OFSAA) environment. This virtually removes the need to write business requirements and re-code the model, which drastically reduces the implementation timeline – given the approaching 2018 deadline for IFRS9 implementation this is a rather appealing feature.

References

Cziráky, D. and Sadleir, P. (2013), Estimation of the Stressed Securitisation Rating Transition Matrices: A Multi-state Markov Approach. *7th International Conference on Computation and Financial Econometrics*.

Gentleman, R., Lawless, J., and Yan, P. (1994), Multi-state Markov models for analyzing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine*, **13**, 805-821.

Jackson, C.H., Sharples, L.D., Thompson, S.G., Du_y, S.W., and Couto, E. (2003), Multistate Markov Models for Disease Progression with Classification Error. *The Statistician*, b(2),1-17.

Jackson, C.H. (2011), Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software*, **38**(8), 1-28.

Kalbeisch, J. and Lawless J. (1985), The Analysis of Panel Data under a Markov Assumption. *Journal of the American Statistical Association*, **80**(392), 863-865.

The R Project for Statistical Computing http://www.r-project.org

Oracle R Entreprise http://www.oracle.com/technetwork/database/database-technologies/r/r-enterprise/overview/index.html

.

ORACLE

Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries
Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US

blogs.oracle.com/oracle
facebook.com/oracle
twitter.com/oracle
oracle.com

Integrated Cloud Applications & Platform Services

MULTI-STATE MARKOV MODELING OF IFRS9 DEFAULT PROBABILITY TERM STRUCTURE IN OFSAA
Authors: Dario Cziráky, PricewaterhouseCoopers (Dario.Cziraky@pwc.com) and Ulrich Zink, Oracle (Ulrich.Zink@oracle.com)