Additional comments on Joe Breeden's papers on Age-Vintage-Time Modeling

Alan Forrest 5/8/25

# 1 Further thoughts

It is now over 20 years since Joe Breeden started writing on Dual Time Dynamics (DTD) and its practical application. He used this approach first to analyse the Asian Economic Crisis in 97-98 for Citibank, but his interest in wine and other scientific applications such dendrochronology, shaped further the method and coined terminology such as "Vintage Analysis"[B 2007]. With improved understanding and better model fitting tools DTD matured into the Age-Vintage-Time (AVT) method that has since been used to provide deep insight to credit portfolio stress testing [B 2016]. In further papers e.g.[B 2016b] and most recently in his newly published book [B 2025], Joe has shown this AVT method can be extended to many more cases in finance and risk, indeed it becomes a core principle in the modern quantitative financial management. We in Credit Risk modeling are fortunate therefore to have Joe's foundational work and case studies on AVT: a coherent method that brings together correctly and powerfully the different time-scales and influences on credit risk and loss.

The preprints linked in the Depository [B 2016] [B 2016b] [BT] record the steps of this journey from foundations to practical application, preparing the modeler to understand and calibrate AVT models for credit management. The papers also make clear the underlying structure and their interpretation, while giving much needed precision to the disambiguation required (specification error) and the staging of model calibration. A future edition of the depository will introduce the more recent advanced applications of AVT.

I have almost nothing to add to such a well-written and interesting development, but for one comment on an early view ([BT] Section 4) that

> No universal solution exists for the model specification error [i.e ambiguity in assignment of linear time trend] described here [i.e. in the AVT model].

It is certainly true that to specificy each of the three terms of the AVT model we need to make choices about model structure, in particular to which term we assign common factors such as linear time trend. [BT] works hard and successfully to resolve these choices for credit risk modelling.

However, in this note I wish to limit this broad statement, by describing how the AVT model outcome (i.e. a probability for each allowed combination of inputs) can be produced in a way that does not need to make explicit each term in the model. By not requiring a specification for each model term, any ambiguity is avoided or at least postponed until we wish to explore further.

But also let me concede immediately that this note's one-stop construction, though practically feasible, is no threat to the approach of [BT]. In Credit Risk each term of the AVT model is separately important and interesting, and the staged approach of [BT] retains overwhelming advantages of control, interpretability and transparency. However, are there contexts where we're not so interested in such intermediate terms, where we might prefer to fit a constrained regression with ambiguous designs by this one-stop approach?

### 1.1 AVT = APC. A note on terminology

The context for the modeling described in Joe's papers and in this note is similar to survival analysis. A data subject (person, account) is observed over a period of time, with a start time (usually account opening) followed by periodic observations up to a last observation which may be an event of interest (uncensored observation), or is merely the latest value we have (censored). In this set up, we have three ways to describe time's influence on a subject's data and actions, and this multiplicity is the essence of the modeling problem to be solved here. It may appear confusing, but the beauty of the AVT modeling approach is to resolve this clearly while losing no information.

Unfortunately the literature contains alterative terms for each of the time scales involved and so causes a further artificial confusion. I am aware of the following choices, and possibly others exist in the fractured, inventive world of Machine Learning.

- Time = Period : The calendar time or time-stamp of an event or data item. Exogenous information, such as econometrics or market factors, enters the AVT model on this time scale. Thus time-series techniques are used typically to fit the time/period term and through it, Stress Testing and other scenario analysis can be implemented.
- Vintage = Cohort : Each data subject has a start time, when they first become ät risk", usually the account start date. Vintage or Cohort is the calendar time of that start, fixed for each subject. Characteristics here are correlated with the customer's and the bank's appetite for risk at that time. Models built on this time-scale are typically static scorecards used for customer management, often built by logistic regression.
- Age = Duration : The length of time since the start time. This is the time scale that captures endogenous effects that evolve during the period of observation, such as customer behaviour changing between the start of the loan and later, or scheduled changes in product design. Models built to this time scale are typically survival analysis or other failure time models.

Thus Age-Vintage-Time and Age-Period-Cohort are interchangeable terms for the kind of modeling we're interested in here, but both are used frequently in the literature. I use Age-Vintage-Time (AVT) throughout this note.

## 2 AVT as a model with structural zeros

AVT is special because it has three time variables as input factors, $a, t, v$, but it does not allow every combination of their values: by definition these three numbers are constrained to two dimensions by a relationship (after rebasing) $a+v=t$. This is an unusual structure for a regression model but it is not unprecedented.

A model with categorical input factors sometimes will exclude certain combinations of attributes as impossible. For example Biological Sex (M/F) + Pregnant (Y/N) disallows MY. Such logical gaps are called **structural zeros** and should be removed in the analysis of the data and in the model build, either by ignoring these combinations and adopting a method that is flexible enough to work with irregular data structures or by fixing a missing value or zero in appropriate cells of the cuboidal table.

I'll build a bridge between AVT and structural zeros, helped by assuming all the times are measured as integer values. This is realistic in credit risk portfolio management where time is usually measured monthly, perhaps daily at most. We still have the constraint $a+v=t$, and so the contingency table of cells indexed by $(a,t,v) \in \mathbb{Z}^3$ has in fact many structural zeros. The non-zero cells of this table are found only indexed by $\Lambda = \{(a,v,t) \in \mathbb{Z}^3 : a+v=t\}$.

The AVT model $p = \mathbb{P}(Default|X,a,v,t)$ uses all three time variables as factors, as well as other factors $X$, and has a structure $\log(p/1-p) = F(a|X)+G(v|X)+H(t|X)$. This takes its form from classic regression which would combine 3 discrete factors without interaction (relative to $X$) using exactly this formula. This formula admits ambiguity in $F,G,H$, for instance in setting the intercept, but in classic regression this is resolved directly by agreeing the model's design vectors. For AVT, such a disambiguating design is more complicated to set, thanks to the presence of time trends and the constraining relation between the factors.

The paper by Breeden and Thomas [BT] resolves completely the ambiguities in choosing $F,G,H$ for ATV in practice. They build the model up in stages: First the calendar time trend is estimated and a detrended $H'$ is calibrated using time-series techniques; then detrended $F'$ is calibrated as a survival analysis including $H'$ as an Offset; then detrended $G'$ is calibrated as a scorecard / logistic regression, with $F'+H'$ as an Offset. This process is iterated and the fit converges on a stable assignment of trend and other common factors among the three terms. The full model then recombines: trend$+F'+G'+H'$. This staging and the order of stages have a practical logic, a clearly explainable structure and each stage's model plays to the strengths of its data and its interpretation.

Never-the-less could the ATV model be fitted all in one shot, and might that make it more simple to derive or at least avoid the ambiguities? The following two constructions combine to show that this can be done for discrete time; and the obstruction to extending to the continuous case seems to be one only of mathematical technique.

## 2.1 Logistic regression and design subspaces

Start with a finite table whose cells are indexed by $\omega \in \Omega$. This table is populated with count data $d = d(\omega)$, a list of positive integers. We think of $d$ as a member of $\mathbb{R}^{+\Omega}$.

Suppose now we choose from the table margins a binary outcome variable, and a selection of input factors possibly with their interaction - i.e. we select a design for a logistic regression model. From this design (independent of the data) we can define a vector subspace $U$ of $\mathbb{R}^{\Omega}$

which allows us to express the calibration of the logistic regression as a constrained convex optimization problem. Find $m$ which solves

$$\operatorname{argmin} \sum_{\omega \in \Omega} m(\omega)(\log m(\omega) - 1) \quad : m \in (d+U) \cap \mathbb{R}^{+\Omega}$$

(with convention $0 \log 0 = 0$). Then $m(\omega)$ gives the cell counts expected by the logistic regression model with this design, fitted to the data $d$, based on the same underlying input population as $d$. Note that in this process we have found $m$ and ensured that it has the expected design, but without determining any of the model parameters.

In fact $U$ can be defined explicitly in terms of the model's design vectors - it is the space orthogonal to all model factor design and population design vectors. In principle this could be used to fit logistic regression models, but better lower dimensional optimisations are used in practice. For this discussion however it is enough to recognise a correspondence between the design and the subspace $U$; and that this produces the model outcomes (on the development population) $m(\omega)$ without having to find any of the model parameters.

## 2.2 Hexagons and design subspaces for AVT models

In AVT we seek appropriate design vectors, i.e. subspace $U$, just as we do for ordinary regression, but the definition of such vectors is complicated materially by the dependence between $a, v, t$. This section details how to accommodate this complication.

First, to keep mathematical issues to a minimum, we restrict the analysis to a finite set of time points by assuming a universal bound on all times by some large value $T$, as would be natural in real conditions. So we set $\Lambda = \{(a, v, t) \in \mathbb{Z}^3 : a + v = t \text{ and } |a|, |v|, |t| \leq T\}$.

To express the designs of interest, we consider certain hexagonal patterns of points as follows.

Start with a basic hexagon, an ordered list $B$ of 6 points in $\Lambda$ as follows:

$$B = (1,0,-1),\ (1,-1,0),\ (0,-1,1),\ (-1,0,1),\ (-1,1,0),\ (0,1,-1)$$

These track round the circumference of a regular hexagon centred at 0. To each of these points we attach a corresponding **alternating weight** $w = +1, -1, +1, -1, +1, -1$.

A **unit hexagon** is a translation of $B$ by $u \in \Lambda : u + B \subset \Lambda$. More generally, a **regular hexagon** is a translation and expansion of $B$ by a positive integer scale $r \in \mathbb{Z}^+ : u + rB \subset \Lambda$.

Given a regular hexagon $C = u + rB \subset \Lambda$ and a real valued function $\phi : \Lambda \to \mathbb{R}$, we can define $S_C\phi = \sum_{b \in B} w(b)\phi(rb + u)$ - the **alternating hexagon sum** for $C$.

It is a nice combinatorial exercise to show that for $\phi : \Lambda \to \mathbb{R}$, the following are equivalent

1. $\phi(a, t, v) = f(a) + g(v) + h(t)$ for some real functions $f, g, h : \mathbb{Z} \to \mathbb{R}$.

2. $S_C\phi = 0$ for all regular hexagons $C$ in $\Lambda$

3. $S_C\phi = 0$ for all unit hexagons $C$ in $\Lambda$

Now consider the vector space $\mathbb{R}^\Lambda$ of real functions $\Lambda \to \mathbb{R}$. For each regular hexagon $C = u + rB$, define the function $s_C : \Lambda \to \mathbb{R}$: $s_C(rb+u) = w(b)$ if $b \in B$ and 0 otherwise. Now define $U$ to be the vector subspace of $\mathbb{R}^\Lambda$ spanned by $s_C$, where $C$ runs over all regular hexagons. Note that by the equivalence of 2 and 3 above, unit hexagons are sufficient to span $U$. So we have a convenient description of a space $U$ that identifies the functional forms of the AVT model. Thus, by the construction of 2.1, we have fitted an AVT model, ensured to have the correct 3 term additive structure, without having to define any of these terms. Thus the ambiguities of AVT specification are avoided.

Note that the discussion above focuses only on the three time variables. The complete AVT model requires attention to other factors $X$ and to the binary outcome as well. However, the design problems for these other marginals are the same problems solved in more standard regressions. This space $U$ provides the essential new idea to solve the unfamiliar problem of calibrating the model to the three time factors, and the full solution for AVT follows a well-known path from here.

Is this any better than the staged approach of [BT]? I think this one-step method will be preferred only in problems where we are not interested in a separate interpretation of each of the model's terms, so that the model built in one step tells us all we want to know. But how different is the case of credit risk!, where each component has its own data structures, interpretation and model method.

So in credit risk there is not much motivation for this one-step calibration method. But it is good to know that, despite its new kinds of complexity, the AVT model conforms to a general regression structure (at least for discrete time) and has a mathematically canonical description that avoids the ambiguities and constraints.

### 2.3 References

[B 2002] Breeden, J.L., 2002. Becoming a better vintner. RMA J. 9, 25–31.

[B 2007] Joseph L. Breeden, Modeling data with multiple time dimensions. Computational Statistics & Data Analysis, Volume 51, Issue 9, 2007, Pages 4761-4785. https://doi.org/10.1016/j.csda.2007.01.023. (https://www.sciencedirect.com/science/article/pii/S0167947307000631)

[B 2016] Breeden, J. L. (2016). Incorporating lifecycle and environment in loan-level forecasts and stress tests. Pre-print June 2016 - see also journal article Incorporating lifecycle and environment in loan-level forecasts and stress tests. European Journal of Operational Research, 255(2), 649–658. https://doi.org/10.1016/J.EJOR.2016.06.008

[B 2016b] Joseph L. Breeden. Forecasting Behavior with Age-Period-Cohort Models: How APC Predicted the US Mortgage Crisis, but Also Does So Much More. Pre-print Paper 2700-2016 for SAS series. March 2016

[BT] Breeden, J., Thomas, L. , Specification errors in stress testing models. Pre-print August 2009 - see also journal article Solutions to specification errors in stress testing models. J Oper Res Soc 67, 830–840 (2016). https://doi.org/10.1057/jors.2015.97

[B 2025] Joseph L. Breeden , Redesigning Credit Risk Modeling to Achieve Profit and Volatility Targets. Bookbaby 2025 ISBN: 9798350978216

Alan Forrest, Edinburgh, August 2025