The Market of Learning

Stochastic Differential Equation Models for Student Growth

Krishna Saxena

June 13, 2024

Abstract

The introduction of technology into the classroom has opened novel avenues for for teachers to improve classroom management and personalized student interventions. This paper explores the application of Stochastic Differential Equation (SDE) based models to fit high-frequency (three observations per week) student achievement data for 14 classrooms. Two paradigms for how students learn are discussed and then translated into SDE models: Random Walk (RW) and Geometric Brownian Motion (GBM). These models are extended to consider student to student interactions and the effects of covariate variables. The RW models are found to have superior performance on the test data, with the RW+Dependent Model attaining a Test Mean Absolute Error of 0.042. The study concludes with discussions about implications of this result on the theory of student learning and practical applications of the model's parameters.

Introduction

Traditionally, quantitative studies of student achievement have relied on end of year assessments. For example, D. D. Goldhaber, Brewer, and Anderson 1999 analyzed exam scores from The National Education Longitudinal Study of 1988. Goldhaber *et al.* employed heirarchical linear models to study the importance of covaraites at different levels of instructional grouping (i.e., school, classroom, within-classroom (individual)) and decompose the variance in test scores attributable to these groupings.

Y. Zhang and L. Zhang 2002 extended this approach over time by analyzing scores from 3rd grade students in 1998 and 2000. Zhang and Zhang found descriptive (but no causal) evidence that there was a shift in the distribution of variance explained at these three grouping levels. Zhang and Zhang's study shows how longitudinal data anlayses have the potential to reveal trends of interest to educational leaders.

Standardized exams are not the only variable of interest for educational researchers. For example, D. Goldhaber and Goodman Young 2024 matched end-of-semester course grades with exam scores for K-12 students across seven academic years. This study design allowed Goldhaber and Goodman Young to describe the weakening relationship between demonstrated achievement on exams and course grade that occurred alongside relaxed grading policies during the COVID-19 pandemic.

As the complexity of longitudinal methodology as grown, so too have the tools to collect student data. Online education platforms such as Khan Academy and IXL allow instructors to collect statistics on each activity their students attempt. These learning platforms offer dashboards so that teachers can aggregate these statistics and adapt their lessons in real-time. The present study's author has used both of the aforementioned websites' application programming interfaces (APIs) as a data scientist intern at The Math Agency (an intervention program for elementary math students).

High-frequency achievement data allows The Math Agency to test theories of student learning. For example, a 2022 project by The Math Agency formalizes measures for practice quantity and quality (Preiner 2022). Specifically, Preiner (2022) defines quantity is the cumulative minutes on learning platforms and quality is overall growth divided by cumulative minutes practiced and used observations of these quantities to discover that "an extra 60 minutes of practice per week corresponds to an extra 1.6 grades of growth over the course of a school year." Additionally, these metrics are shared with teachers who can then spend time understanding how to motivate the students with the lowest practice quantity and teach tailored strategies to the students who are growing the lowest despite their best efforts.

Such applications of high-frequency growth data to concrete pedagogical interventions highlight the promise of educational data analysis. In 2023, this study's author created a tool where teachers could document academic interventions and choose metrics to be tracked in association with each intervention. Using piecewise linear regressions and bootstrapping, the author's tool would let instructors know whether the intervention was specific, allowing the team to measure the relative success rates of intervention strategies (see pages 1-3 of The Math Agency's 2023-24 Social Purpose Report (Math Agency 2024) for more details).

More accurate models for student achievement can help educators answer essential questions about their students. For example, in a 2023 article answering the pressing question "Can we repeatably close education gaps?" Preiner uses linear models to show how much continuous suport a typical student at three partner schools require to reach grade-level proficiency. Such analyses are essential to understand whether the organization is reaching its mission of closing education gaps, but lack the detail a teacher needs to inform his/her day-to-day interactions with students who come into the classroom with different preparation levels and growth trajectories. *The current* *study expands the literature by applying Stochastic Differential Equation (SDE) based models to high-frequency education forecasting, with an emphasis on disussing how final model parameters can inform classroom management.*

Mathematical Background

While the author presumes that readers have an understanding of stochastic calculus equivalent to a student who has completed MATH/STAT 493, a brief introduction of the field's foundational results are included.

2.1 Brownian Motion

The author would like to note that S.E. Shreve provides detailed proofs of the properties discussed below in Chapter 3 of Volume II of his book *Stochastic Calculus for Finance* Shreve 2004.

A Brownian Motion (BM) is a continuous stochastic process, that is, it evolves in continuous time and its evolution is governed by a probability distribution. $\{B_t \;\; t \geq 0\}$ is BM if:

a. $B_0 = b_0 \;\; b_0 \in [R]$. Generall, by convention, and in every BM for this paper, $b_0 := 0$

b. $\forall 0 = t_0 < t_1 < \ldots < t_n, B_{t_{i+1}} - B_{t_i} \perp B_{t_i} - B_{t_{i-1}}$ and $B_t - B_s \sim \mathcal{N}(0, t-s)$. This property will be referred to as independent, Normal increments.

c. The map $t \to W_t$ is continuous almost surely.

These properties also imply that BM is a Martingale:

$$B_{t+\Delta t} \;\; | \, \sigma(B_s \;\, : s \leq \;\, t) \sim \;\, N(B_t, \;\, \Delta t)$$

2.2 Itô's Rule

There are two essential properties for the derivatives of a single BM

a. $dB_t \cdot dB_t = dt$

b. $\forall n \geq 3, (dB_t)^n = 0 \Rightarrow \forall n \geq 1, dt(dB_t)^n = 0$

Itô's Rule can be naturally extended to multiple dimensions. Suppose BM W_t that moves independently of B_t, then

$$dB_t \cdot dW_t = 0$$

Stochastic Processes can always be made by combining BM, e.g. let's construct X_t, Y_t as a preview of Covariance anlaysis that will come later.

$$X_t = \alpha B_t + \beta W_t$$
$$Y_t = \gamma B_t + \delta W_t$$

2.3 Itô's Formula

In a one-dimensional setting, Itô's Formula Itô 1944, also called "The Fundamental Theorem of Stochastic Calculus" is as follows: Suppose $f \in C_2(\mathbb{R}) \to \mathbb{R}$, then

$$f(X_{t+\Delta t}) - f(X_t) = \int_t^{t+\Delta t} f'(X_s)dX_s + \frac{1}{2}\int_t^{t+\Delta t} f''(X_s)(dX_s)^2 \equiv$$

an equivalent differential form

$$df(X_t) = f'(X_t)dX_t + \frac{f''(X_t)}{2}(dX_t)^2$$

In two dimensions, we can let $f \in C_{2,2}(\mathbb{R}^2) \to \mathbb{R}$, then

$$\begin{aligned}
f(X_{t+\Delta t}, Y_{t+\Delta t}) - f(X_t, Y_t) &= \int_t^{t+\Delta t} f_x(X_s, Y_s)dX_s + \int_t^{t+\Delta t} f_y(X_s, Y_s)dY_s \\
&+ \frac{1}{2}\int_t^{t+\Delta t} f_{xx}(X_s, Y_s)(dX_s)^2 \\
&+ 2\cdot\frac{1}{2}\int_t^{t+\Delta t} f_{xy}(X_s, Y_s)(dX_s)(dY_s) \\
&+ \frac{1}{2}\int_t^{t+\Delta t} f_{yy}(X_s, Y_s)(dY_s)^2 \equiv \\
df(X_t, Y_t) &= f_x(X_t, Y_t)dX_t + f_y(X_t, Y_t)dY_t + \frac{f_{xx}(X_t, Y_t)}{2}(dX_t)^2 \\
&+ f_{xy}(X_t, Y_t)(dX_t)(dY_t) \\
&+ \frac{f_{yy}(X_t, Y_t)}{2}(dY_t)^2
\end{aligned}$$

2.4 Itô-Doeblin Formula

When the function explicitly depends on t, then there are extra assumptions on f and a new form for its difference formula Itô 1951.

Suppose $f \in C_{1,2}(\mathbb{R}^+, \mathbb{R}) \to \mathbb{R}$, then

$$\begin{aligned}
f(t+\Delta t, X_{t+\Delta t}) - f(t, X_t) &= \int_t^{t+\Delta t} f_t(s, X_s)ds + \int_t^{t+\Delta t} f_x(s, X_s)dX_s + \frac{1}{2}\int_t^{t+\Delta t} f_{tt}(s, X_s)(ds)^2 \\
&+ 2\cdot\frac{1}{2}\int_t^{t+\Delta t} f_{tx}(s, X_s)(dsdX_s) + \frac{1}{2}\int_t^{t+\Delta t} f_{xx}(s, X_s)(dX_s)^2
\end{aligned}$$

By Itô's Rule $(ds)^2 = (dsdX_s) = 0$ when X_t is stochastic and based on some BM B_t

$$= \int_t^{t+\Delta t} f_t(s, X_s)ds + \int_t^{t+\Delta t} f_x(s, X_s)dX_s + \frac{1}{2}\int_t^{t+\Delta t} f_{xx}(s, X_s)(dX_s)^2 \equiv$$
$$df(t, X_t) = f_t(t, X_t)dt + f_x(t, X_t)dX_t + \frac{f_{xx}(t, X_t)}{2}(dX_t)^2$$

If a function $f \in C_{1,2,2}(\mathbb{R}^+, \mathbb{R}, \mathbb{R}) \to \mathbb{R}$ depends on mulitple random variables, then (ignoring cross terms $f_{t\_}$ as as the fact that $dt \cdot d\_ = 0$ because of Itô's Rule has been explored previously)

$$\begin{aligned}
f(t + \Delta t, X_{t+\Delta t}, Y_{t+\Delta t}) - f(t, X_t, Y_t) &= \int_t^{t+\Delta t} f_t(s, X_s, Y_s)ds + \int_t^{t+\Delta t} f_x(s, X_s, Y_s)dX_s + \int_t^{t+\Delta t} f_y(s, X_s, Y_s)dY_s \\
&+ \frac{1}{2}\int_t^{t+\Delta t} f_{xx}(s, X_s, Y_s)(dX_s)^2 \\
&+ 2 \cdot \frac{1}{2}\int_t^{t+\Delta t} f_{xy}(s, X_s, Y_s)(dX_s)(dY_s) \\
&+ \frac{1}{2}\int_t^{t+\Delta t} f_{yy}(s, B_s, W_s)(dY_s)^2 \equiv \\
df(t, X_t, Y_t) &= f_t(t, X_t, Y_t)dt + f_x(t, X_t, Y_t)dX_t + f_y(t, X_t, Y_t)dY_t \\
&+ \frac{f_{xx}(t, X_t, Y_t)}{2}(dX_t)^2 + f_{xy}(t, X_t, Y_t)(dX_t)(dY_t) \\
&+ \frac{f_{yy}(s, B_s, W_s)}{2}(dY_s)^2
\end{aligned}$$

Theories for Student Growth and Their SDE Models

This paper investigates two models for how students learn. Please note that the author does not claim to show that either of these is accurate, rather this study can provide descriptive evidence that one model is a better fit for the students in the dataset. The studied models are intended to have utility for teachers while planning instruction, not give value judgments on a student's aptitude. Here is a summary of the idea behind these models:

a. **Arithmetic Growth Model**: students learn a fixed amount daily, plus or minus noise based on external circumstances. In Stochastic Calculus literature, such models are called Random Walks.

b. **Geometric Growth Model**: students grow there knoweldge by a fixed fraction daily, this fraction fluctuates noisily based on external circumstances. In Stochastic Calculus literature, this model is called Geometric Brownian Motion.

3.1 Arithmetic Growth

3.1.1 Building Intuition

To understand the Arithmetic Growth model, it is instructive to first consider a discrete timeline. Imagine that at time t, a student has achieved S_t units of knowledge for some continuous measure of learning. Then, at the next timestep 1 unit of time later, the student learns u units with probability p and $d < u$ units with probability $1-p$, independently of the movement up to t:

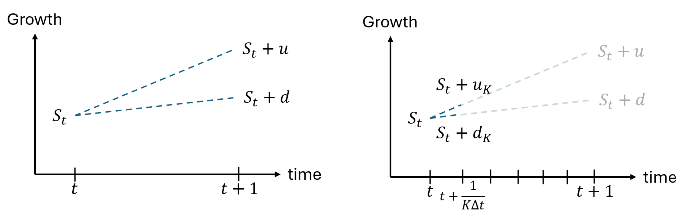
$$S_{t+1} = S_t + \begin{cases} u & w.p.\ p \\ d & w.p.\ 1-p \end{cases}$$

Ideally, Arithmetic Growth models will satisfy $\mathbb{E}\left[S_{t+\Delta t} - S_t\right] = \mu\Delta t, \mathrm{Var}\left[S_{t+\Delta t} - S_t\right] = \sigma^2\Delta t$ so that they can be modeled with BM B_t: $S_{t+\Delta t} = S_t + \mu\Delta t + \sigma\left(B_{t+\Delta t} - B_t\right)$.

To find such μ, σ, it is helpful to cut the time interval $[t, t+\Delta t]$ into $K\Delta t$ intervals of size $\frac{1}{K}$ time units (for arbitrary $K, \Delta t \in \mathbb{N}$). Holding p constant, the values u_K, d_K will scale according to K such that $u_K, d_K \to 0$ as $K \to \infty$. Then,

$$S_{t+\frac{1}{K\Delta t}} = S_t + \begin{cases} \frac{u_K}{} & w.p.\ p \\ \frac{d_K}{} & w.p.\ 1-p \end{cases},$$

$$S_{t+\Delta t} = S_t + (K\Delta t)d_K + \sum_{i=1}^{K\Delta t}(u_K - d_K)\cdot b_i : \{b_i\}_{i=1}^{K} \overset{\text{i.i.d.}}{\sim} \mathrm{Bern}(p)$$



Visualizations of discrete-time versions of the Arithmetic Growth Model

Under this setting,

$$\mathbb{E}\left[S_{t+\Delta t} - S_t\right] = \mathbb{E}\left[(K\Delta t)d_K + \sum_{i=1}^{K\Delta t}(u_K - d_K)\cdot b_i\right]$$

By linearity of expectation, $\because$ all b_i have indentical distributions

$$\begin{aligned} &= (K\Delta t)d_K + (u_K - d_K)\sum_{i=1}^{K\Delta t}\mathbb{E}\left[b_i\right] \\ &= (K\Delta t)(d_K + p(u_K - d_K)) \\ &= (K\Delta t)(p\cdot u_K + (1-p)d_K), \end{aligned}$$

$$\begin{aligned} \mathrm{Var}\left[S_{t+\Delta t} - S_t\right] &= \mathrm{Var}\left[(K\Delta t)d_K + \sum_{i=1}^{K\Delta t}(u_K - d_K)\cdot b_i\right] \\ &= \mathrm{Var}\left[\sum_{i=1}^{K\Delta t}(u_K - d_K)\cdot b_i\right] \end{aligned}$$

$\because$ all b_i are i.i.d.

$$= (K\Delta t)(u_K - d_K)^2 \operatorname{Var}[b_i]$$
$$= (K\Delta t)(u_K - d_K)^2 p(1-p)$$

Thus, a $S_{t+\Delta t} = S_t + \mu\Delta t + \sigma(B_{t+\Delta t} - B_t)$ where B_t is BM would be a good fit for an **additive model of learning**.

3.1.2 SDE Models for the Arithmetic Growth Model

This BM-based model is now studied in detail. The following derivations are studied conditioned on the sigma field $\{\sigma(B_s)\}_{s=0}^{t}$, i.e., any randomness caused by fluctuations of BM B_s up to and including time t are understood to be constant.

$$S_{t+\Delta t} - S_t = \mu\Delta t + \sigma(B_{t+\Delta t} - B_t)$$

Let $f(t,x) = \mu t + \sigma x$. Using The Itô-Doeblin Formula

$$\begin{aligned} dS_t &= f_t(t, B_t)dt + f_x(t, B_t)dB_t \\ &= \mu dt + \sigma dB_t \end{aligned} \tag{1}$$

To model multiple students in a classroom, the SDE in (1) can be vectorized to $\mathbb{R}^{N_{stu}}$ such that $N_{stu} :=$ the number of students in a classroom, $S_{i,t}$ is the i^{th} student's growth timeseries, μ_i is the i^{th} student's growth rate, and σ_i is the volatility in the i^{th} student's growth rate. Because this multivariate SDE is simply the concatenation of N_{stu} univariate SDEs, it can be integrated in the same manner. This model will be referred to as the **Independent Arithmetic Growth Model**:

$$\begin{aligned} d\vec{S}_t &= \vec{\mu}dt + \vec{\sigma}d\vec{B}_t, \\ \vec{S_{t+\Delta t}} &= \vec{S}_t + \vec{\mu}\Delta t + \vec{\sigma}\left(\vec{B_{t+\Delta t}} - \vec{B}_t\right) \end{aligned}$$

$\because$ BM has independent, Normal increments $\forall i \in N_{stu}, B_{i,t+\Delta t} - B_{i,t} \sim \mathcal{N}(0, \Delta t)$

$$= \vec{S}_t + \vec{\mu}\Delta t + \vec{\sigma}\mathcal{N}(\vec{0}, \Delta t I)$$

A more interesting, and useful, model accounts for interactions between students. We make the assumption that students have neither the ability to directly change each other's knowledge levels at any timestep nor the ability to change each other's overall growth rates $\vec{\mu}$. Rather, students can only cause fluctuations in each other's growth during a specific time interval (e.g., by sharing a strategy with a classmate or distracting a friend). Such phenomena can be modeled by the SDE

$$d\vec{S}_t = \vec{\mu}dt + \Sigma d\vec{B}_t \tag{2}$$

where $\Sigma \in \mathbb{R}^{N_{stu} \times N_{stu}}$ is a Positive Definite (PD) matrix and $\Sigma_{ij} := \rho_{ij}\sigma_i\sigma_j$ where ρ_{ij} is the covariance between student i and j and σ_i and σ_j are their respective student's volatility (i.e., a consistenely focused student working on content at the right level will have low volatility, but distractions or poorly targetted lessons will cause fluctuations in a student's growth pattern). This model will be referred to as the **Dependent Arithmetic Growth Model**

To integrate the SDE (2), it is constructive to look at a single student at a time. For arbitrary student i, $d\S_{i,t} = \mu_i dt + \Sigma_{i,:} d\vec{B}_t$ where $\Sigma_{i,:}$ denotes the i^{th} row of Σ.

$$\begin{aligned} dS_{i,t} &= \mu_i dt + \Sigma_{i,:} d\vec{B}_t \\ &= \mu_i dt + \sum_{j=1}^{N_{stu}} \Sigma_{i,j} dB_{j,t} \Rightarrow \\ \int_{s=t}^{t+\Delta t} dS_{i,s} &= \int_{s=t}^{t+\Delta t} \mu_i ds + \int_{s=t}^{t+\Delta t} \sum_{j=1}^{N_{stu}} \Sigma_{i,j} dB_{j,s} \end{aligned}$$

$\because$ sums and integrals are linear operators

$$\begin{aligned} &= \mu_i s_t^{t+\Delta t} + \sum_{j=1}^{N_{stu}} \Sigma_{i,j} \int_{s=t}^{t+\Delta t} dB_{j,s} \\ &= \mu_i \Delta t + \sum_{j=1}^{N_{stu}} \Sigma_{i,j} dB_{j,s}{}_{s=t}^{t+\Delta t} \end{aligned}$$

$\because$ BM has independent, Normal increments $\forall i \in N_{stu}, B_{i,t+\Delta t} - B_{i,t} \sim \mathcal{N}(0, \Delta t)$

$$\begin{aligned} &= \mu_i \Delta t + \sum_{j=1}^{N_{stu}} \Sigma_{i,j} \mathcal{N}(\vec{0}, \Delta t) \\ &= \mu_i \Delta t + \Sigma_{i,:} \mathcal{N}(\vec{0}, \Delta t) \Rightarrow \\ S_{i,s}{}_{s=t}^{t+\Delta t} &= \mu_i \Delta t + \Sigma_{i,:} \mathcal{N}(\vec{0}, \Delta t) \Rightarrow \\ S_{i,t+\Delta t} &\sim S_{i,t} + \mu_i \Delta t + \Sigma_{i,:} \mathcal{N}(\vec{0}, \Delta t) \therefore \\ \vec{S_{t+\Delta t}} &\sim \vec{S}_t \vec{\mu} \Delta t + \Sigma \mathcal{N}(\vec{0}, \Delta t I) \end{aligned} \tag{3}$$

The third and final Arithmetic Growth Model of interest is the **Dependent + Covariate Arithmetic Growth Model**. This model differs from the **Dependent Arithmetic Growth Model** by the addition of covariate data to the growth term. The **Dependent + Covariate Arithmetic Growth Model** will be defined as the SDE:

$$d\vec{S}_t = \left(\vec{\mu} + C_t \vec{\theta}\right) dt + \Sigma d\vec{B}_t$$

where $C_t \in \mathbb{R}^{N_{stu} \times N_{covs}}, \vec{\theta} \in \mathbb{R}^{N_{covs}}$ and N_{covs} is the number of covariate variables in a model specification. Note that the previous equation is equivalent to

$$d\vec{S}_t = \begin{bmatrix} I & C_t \end{bmatrix} \begin{bmatrix} \vec{\mu} \\ \vec{\theta} \end{bmatrix} dt + \Sigma d\vec{B}_t \tag{4}$$

This SDE can be integrated in the same way:

$$\vec{S_{t+\Delta t}} \sim \vec{S}_t + \begin{bmatrix} I & C_t \end{bmatrix} \begin{bmatrix} \vec{\mu} \\ \vec{\theta} \end{bmatrix} \Delta t + \Sigma \mathcal{N}(\vec{0}, \Delta t) \tag{5}$$

3.2 Geometric Growth Model

3.2.1 Building Intuition

A Geometric Brownian Motion model for the growth of a financial asset's price was studied extensively by R.C. Merton Merton 1973 in his statistical analysis of F. Black and M. Scholes' option pricing formula Black and Scholes 1973. An intuitive explanation of this model can be found in Vol I, Chapter 1 of S.E. Shreve's book Shreve 2004. The following is an adaptation of the latter, in the context of student growth.

Much like the Arithmetic Growth Model, imagine that at time t, a student has achieved S_t units of knowledge for some continuous measure of learning. Then, one unit of time later, the student grows that S_t base knowledge-units by a factor of u with probability p and $d < u$ with probability $1 - p$, independently of the movement up to t:

$$S_{t+1} = S_t \cdot \begin{cases} u^{\Delta t} & w.p.\ \ p \\ d^{\Delta t} & w.p.\ \ 1-p \end{cases}$$

Assume that a student has some averate growth rate $\mu \ :\ \mathbb{E}\left[\ln\left(\frac{S_{t+\Delta t}}{S_t}\right)\right] = \mu\Delta t$ and some mean volatility in growth rate $\sigma \ :\ \mathrm{Var}\left[\frac{S_{t+\Delta t}}{S_t}\right] = \sigma^2\Delta t$ that both scale in Δt. In this scenario, the parameters of this process do not scale evenly as we decrease the time interval between measurements. We observe that after scaling down Δt by some factor $K \in \mathbb{N}$ while holding p constant, $u_K = \sqrt[K]{u}, d_K = \sqrt[K]{d}$ (Note that $u_K, d_K \approx 1$ but $u_K > d_K$):

$$S_{t+\frac{\Delta t}{K}} = S_t \cdot \prod_{j=1}^{K} G_j^K \ \ : \{G_j^K\}_{j=1}^K \overset{\text{i.i.d.}}{\sim} \begin{cases} u^{\frac{\Delta t}{K}} & w.p.\ \ p \\ d^{\frac{\Delta t}{K}} & w.p.\ \ 1-p \end{cases}$$

By the i.i.d. distribution of $\{G_j^K\}_{j=1}^K$ and similar reasoning to the corresponding derivations for the Arithmetic Growth model, it becomes apparent that such μ, σ exist for any given K.

3.2.2 SDE Models for the Geometric Growth Model

Analysis of the **Independent Geometric Growth Model** starts with the same SDE studied in Merton 1973:

$$\begin{aligned} dS_t &= S_t(\mu dt + \sigma dB_t) \Rightarrow \\ \frac{dS_t}{S_t} &= \mu dt + \sigma dB_t \Rightarrow \end{aligned}$$

For $\frac{dS_t}{S_t}$, a good Ansatz is $S_t = f(x) = \ln(x)$. Then, $f'(x) = \frac{1}{x}, f''(x) = \frac{-1}{x^2}$.

Let $f(x) = ln(x)$. Using Itô's Formula:

$$\begin{aligned} df(S_t) &= f'(S_t)dS_t + \frac{1}{2}f''(S_t)(dS_t)^2 \Rightarrow \\ d\ln(S_t) &= \frac{1}{S_t}dS_t - \frac{1}{2S_t^2}(dS_t)^2 \\ &= \frac{1}{S_t}\left(S_t(\mu dt + \sigma dB_t)\right) - \frac{1}{2S_t^2}\left(S_t(\mu dt + \sigma dB_t)\right)^2 \\ &= (\mu dt + \sigma dB_t) - \frac{1}{2}(\mu dt + \sigma dB_t)\cdot(\mu dt + \sigma dB_t) \end{aligned}$$

By Itô's Rule, $dB_t \cdot dB_t = dt, dt \cdot d\_ = 0$

$$= (\mu dt + \sigma dB_t) - \frac{1}{2}\left(\sigma^2 dt + 0\right)$$

$$= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dB_t \Rightarrow$$

$$\int_{s=t}^{t+\Delta t} d\ln(S_s) = \int_{s=t}^{t+\Delta t} \left(\mu - \frac{\sigma^2}{2}\right) ds + \int_{s=t}^{t+\Delta t} \sigma dB_s \Rightarrow$$

$$\ln\left(S_{t+\Delta t}\right) - \ln\left(S_t\right) = \left(\mu - \frac{\sigma^2}{2}\right) \Delta t + \sigma(B_{t+\Delta t} - B_t) \Rightarrow$$

$$S_{t+\Delta t} \sim e^{\left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma \mathcal{N}(0, \Delta t)}$$

The **Dependent Geometric Growth Model** and **Dependent + Covariate Geometric Growth Model**, like their corresponding Arithmetic Growth Models differ in just the drift term. Thus, their SDEs will be studied at the same time:

As before, let $C_t \in \mathbb{R}^{N_{stu} \times N_{covs}}, \vec{\theta} \in \mathbb{R}^{N_{covs}} : N_{covs}$ is the number of covariate variables in a model specification. Then, the **Dependent + Covariate Geometric Growth Model**'s SDE is

$$d\vec{S}_t = S_t \left(\left(\vec{\mu} + C_t\vec{\theta}\right) dt + \Sigma d\vec{B}_t\right)$$

It is again constructive to look at a single student's growth series:

$$\begin{aligned} d\vec{S}_{i,t} &= S_{i,t}\left((\mu_i + C_{t,i,:}\vec{\theta})dt + \Sigma_{i,:} d\vec{B}_t\right) \\ &= S_{i,t}\left((\mu_i + C_{t,i,:}\vec{\theta})dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t}\right) \Rightarrow \end{aligned}$$

Let $f(x) = ln(x)$. Then, $f'(x) = \frac{1}{x}, f''(x) = \frac{-1}{x^2}$. Using Itô's Formula

$$\begin{aligned} df(S_t) &= f'(S_t)dS_t + \frac{1}{2} f''(S_t)(dS_t)^2 \Rightarrow \\ d\ln(S_t) &= \frac{1}{S_t} dS_t - \frac{1}{2S_t^2}(dS_t)^2 \\ &= \frac{1}{S_t} \cdot \left(S_{i,t}\left((\mu_i + C_{t,i,:}\vec{\theta})dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t}\right)\right) - \frac{1}{2S_t^2}\left(S_{i,t}\left((\mu_i + C_{t,i,:}\vec{\theta})dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t}\right)\right)^2 \\ &= \left((\mu_i + C_{t,i,:}\vec{\theta})dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t}\right) - \frac{\left((\mu_i + C_{t,i,:}\vec{\theta})dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t}\right)^2}{2} \end{aligned}$$

By Itô's Rule, $dt \cdot d\_ = 0$

$$\begin{aligned} &= \left((\mu_i + C_{t,i,:}\vec{\theta})dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t}\right) - \frac{\left(\sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t}\right) \cdot \left(\sum_{k=1}^{N_{stu}} \Sigma_{ik} dB_{k,t}\right)}{2} \\ &= \left((\mu_i + C_{t,i,:}\vec{\theta})dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t}\right) - \frac{\left(\sum_{j=1}^{N_{stu}} \Sigma_{ij}^2 dB_{j,t}^2\right) \cdot \left(\sum_{k=1}^{N_{stu}} \Sigma_{ik} dB_{k,t}\right)}{2} \\ &= \left((\mu_i + C_{t,i,:}\vec{\theta})dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t}\right) - \frac{\left(\sum_{j=1}^{N_{stu}} \Sigma_{ij}^2 dB_{j,t}^2\right) + \left(\sum_{1 \le j < k \le N_{stu}} \Sigma_{ik} dB_{k,t}\right)}{2} \end{aligned}$$

By Itô's Rule $dB_{i,t} \cdot dB_{i,t} = dt$. Additionally, by construction each student's BM is i.i.d., i.e., $\forall i \neq j, dB_{i,t} \cdot dB_{j,t} = 0$

$$= \left((\mu_i + C_{t,i,:}\vec{\theta})dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t} \right) - \frac{\sum_{j=1}^{N_{stu}} \Sigma_{ij}^2 dt}{2}$$
$$= \left(\mu_i + C_{t,i,:}\vec{\theta} - \frac{\sum_{j=1}^{N_{stu}} \Sigma_{ij}^2}{2} \right) dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t} \Rightarrow$$

Note that by definition of matrix multiplication, $\sum_{j=1}^{N_{stu}} \Sigma_{ij}^2 = \sum_{j=1}^{N_{stu}} \Sigma_{ij} \cdot \Sigma_{ji}^T = (\Sigma\Sigma^T)_{ii}$

$$= \left(\mu_i + C_{t,i,:}\vec{\theta} - \frac{(\Sigma\Sigma^T)_{ii}}{2} \right) dt + \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t} \Rightarrow$$

$$\int_{s=t}^{t+\Delta t} d\ln(S_{i,s}) = \int_{s=t}^{t+\Delta t} \left(\mu_i + C_{s,i,:}\vec{\theta} - \frac{\sum_{j=1}^{N_{stu}} \Sigma_{ij}^2}{2} \right) ds + \int_{s=t}^{t+\Delta t} \sum_{j=1}^{N_{stu}} \Sigma_{ij} dB_{j,t} \Rightarrow$$

Under the assumption that covariates do not change in the interval $[t, \Delta t]$

$$\ln(S_{i,t+\Delta t}) - \ln(S_{i,t}) = \left(\mu_i + C_{t,i,:}\vec{\theta} - \frac{(\Sigma\Sigma^T)_{ii}}{2} \right) \Delta t + \sum_{j=1}^{N_{stu}} \Sigma_{ij} \left(B_{j,t+\Delta t} - B_{j,t} \right) \Rightarrow$$
$$S_{i,t+\Delta t} \sim S_{i,t} \cdot \exp \left\{ \left(\mu_i + C_{t,i,:}\vec{\theta} - \frac{(\Sigma\Sigma^T)_{ii}}{2} \right) \Delta t + \sum_{j=1}^{N_{stu}} \Sigma_{ij} \mathcal{N}(\vec{0}, \Delta t I) \right\}$$

This expression can be vectorized again:

$$\vec{S_{t+\Delta t}} \sim \vec{S}_t \cdot \exp \left\{ \left(\vec{\mu} + C_t\vec{\theta} - \frac{\mathrm{diag}\,(\Sigma\Sigma^T)}{2} \right) \Delta t + \Sigma \mathcal{N}(\vec{0}, \Delta t I) \right\}$$
$$\equiv \vec{S}_t \cdot \exp \left\{ \begin{bmatrix} I & C_t \end{bmatrix} \begin{bmatrix} \vec{\mu} - \frac{1}{2}\mathrm{diag}\,(\Sigma\Sigma^T) \\ \theta \end{bmatrix} \Delta t + \Sigma \mathcal{N}(\vec{0}, \Delta t I) \right\} \tag{6}$$

Data and Methods

4.1 Dataset

The data used in this study comes from students in The Math Agency's 2023-24 and 2024-25 cohorts. Measurements for these students were taken from internal sources (e.g., attendance data) or their program-issued Khan Academy accounts. The mean measurement of student progress in this study is Khan Academy mastery points scaled to grade-levels learned (to be interpretable and consistent with other public Math Agency analyses).

In mathematical notation, the overall dataset $\mathcal{D}$ consists of N_{co} subdatasets $\{\mathcal{D}_i\}_{i=0}^{N_{co}}$ where $N_{co} = 14$ is the number of Math Agency cohorts with high-frequency Khan Academy practice schedules. Each subdataset $\mathcal{D}_i$ is a timeseries that can be further decomposed:

$$\mathcal{D}_i = \{t, \vec{S}_t, C_t\}_{t=t_0}^{t_T}$$

where $t \in [t_0, t_f]$ are increasing, but inconsistently spaced timestamps when students practice (measured in days), $\vec{S}_t \in \mathbb{R}^{+,N_{stu}}$ is a vector of mtudent mastery at time t (note that N_{stu} changes across cohorts/subdatasets), and $C_t \in \mathbb{R}^{N_{stu} \times N_{covs}}$ are vecotrs of covariate data for each student at time t.

Across the 14 cohorts, the average number of timestamps T is 159.4 (SD = 49.0) and average number of students N_{stu} is 25.0 (SD = 13.3). Because the objective of dependent models was quantifying student-student interactions, students who are missing more than 66% of mastery observations are dropped from the datasets in this analysis. The students who are dropped by this filter either left the program early, or joined late. For the remaining students, any missing signal values were filled by independent aritmetic or geometric models (corresponding to which model was being studied) with their σ parameter scaled down by 10 (to prevent imputated values from biasing the $Sigma$ parameter in future analyses). Covariate data was imputed by forward filling the last non-missing value, then backward filling the last non-missing value, and finally filling with the cohort mean for that variable.

After dropping students with excessive missing growth data, there average of N_{stu} across cohorts is 8.1 (SD = 5.0) students.

Furthermore, to understand how well models generalize to future data, each $\mathcal{D}_i$ is split such that 70%, 20%, and 10% of timestamps form the i^{th} train, validation, and test subsets: ($\mathcal{D}_{train,i}$, $\mathcal{D}_{val,i}$, $\mathcal{D}_{test,i}$, respectively). All i train, validation, and test subsets are then combined to form the train, validation, and test data sets ($\mathcal{D}_{train}$, $\mathcal{D}_{val}$, $\mathcal{D}_{test}$).

4.2 Maximum Likelihood Estimation (MLE)

4.2.1 MLE for Arithmetic Growth Models

Since the **Independent Arithmetic Growth Model** and **Dependent Arithmetic Growth Model** are special instances of the **Dependent + Covariate Arithmetic Growth Model** (i.e., when N_{stu} is forced to 1, and $\vec{\theta}$ is forced to $\vec{0}$, respectively), finding the MLE parameters for (5) is sufficient for finding all aritmetic growth models' MLE parameters.

Becasue the shapes and meaning of parameters $\vec{\mu}, \Sigma$ are dependent on the classroom being modeled, MLE parameters must be found for each subdataset and final global (shared) parameters (i.e., $\vec{\theta}$) will be averaged across subdatasets.

Recall that the i^{th} subdataset is notated $\mathcal{D}_i = \{t, \vec{S}_t, C_t\}_{t=t_0}^{t_T}$. Let j be an aribtrary index of $\mathcal{D}_i$. The random portion of (5) can be isolated to find the probability density of $\vec{S_{t_{j+1}}}$ given $t_{j+1}, t_j, \vec{S_{t_j}}, C_{t_j}$. **Note that the density of the random variable in** (5) **is used to find the probability and not the RV in** (4)**.** If the later was chosen instead, the effects of $\left(d\vec{S}_t\right)^\top \left(d\vec{S}_t\right)$,

and thus Itô's Rule that $dB_t^2 = dt$, would not be modeled:

$$\vec{S_{t_{j+1}}} \sim \vec{S_{t_j}} + \begin{bmatrix} I & C_{t_j} \end{bmatrix} \begin{bmatrix} \vec{\mu} \\ \vec{\theta} \end{bmatrix} (t_{j+1} - t_j) + \Sigma \mathcal{N}(\vec{0}, (t_{j+1} - t_j)\, I) \Rightarrow$$

$$\vec{S_{t_{j+1}}} - \vec{S_{t_j}} - \begin{bmatrix} I & C_{t_j} \end{bmatrix} \begin{bmatrix} \vec{\mu} \\ \vec{\theta} \end{bmatrix} (t_{j+1} - t_j) \sim \Sigma \mathcal{N}(\vec{0}, (t_{j+1} - t_j)\, I) \Rightarrow$$

Let $\Delta t_j := t_{j+1} - t_j,\ \Delta \vec{S}_j := \vec{S_{t_{j+1}}} - \vec{S_{t_j}},\ \vec{\alpha} = \begin{bmatrix} \vec{\mu} \\ \vec{\theta} \end{bmatrix}$

$$\Delta \vec{S}_{t_j} + \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha} \sim \mathcal{N}(\vec{0}, \Delta t_j \Sigma) \tag{7}$$

Applying the MV Normal probability density function to (7), the likelihood of the j^{th} interval of subdataset i under the **Dependent + Covariate Arithmetic Growth Model** is

$$\begin{aligned} f(\mathcal{D}_i j \mid \vec{\mu}, \Sigma, \vec{\theta}) &= (2\pi)^{\frac{-N_{stu}}{2}} \cdot |\Delta t_j \Sigma|^{\frac{-1}{2}} \cdot \exp\left\{ \frac{-1}{2} \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha} \right)^\top (\Delta t_j \Sigma)^{-1} \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha} \right) \right\} \\ &= (2\pi)^{\frac{-N_{stu}}{2}} \cdot |\Delta t_j \Sigma|^{\frac{-1}{2}} \cdot \exp\left\{ \frac{-1}{2\Delta t_j} \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha} \right)^\top \Sigma^{-1} \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha} \right) \right\} \Rightarrow \end{aligned}$$

The log –likelihood of the j^{th} interval is

$$\ell(\mathcal{D}_{ij} \mid \vec{\mu}, \Sigma, \vec{\theta}) = -\frac{N_{stu}}{2} \ln(2\pi) - \frac{\ln(\Delta t_j) + \ln(|\Sigma|)}{2}$$

Because BM is the only modeled source of randomness in the evolution of $\vec{S}_t$ and BM has an independent increments property of BM, the likelihood of the timeseries is the product of the likelihood of its intervals. Thus, the average log –likelihood of the i^{th} subdataset is

$$\begin{aligned} \bar{\ell}(\mathcal{D}_{ij} \mid \vec{\mu}, \Sigma, \vec{\theta}) = &-\frac{N_{stu}}{2} \ln(2\pi) - \frac{\sum_{j=0}^{T} \ln(\Delta t_j)}{2T} - \frac{\ln(|\Sigma|)}{2} \\ &- \frac{1}{2T} \sum_{j=0}^{T-1} \frac{\left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha} \right)^\top \Sigma^{-1} \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha} \right)}{\Delta t_j} \end{aligned} \tag{8}$$

The MLE values for $\hat{\alpha}$ (which can then be separated to $\hat{\mu}, \hat{\theta}$) and $\hat{\Sigma}$ can are the roots of the gradient of 8. Noting the multivariable chain rule $\frac{df \circ \vec{g}(X)}{dX} |_x = \vec{\nabla}_g (X)^\top |_x\ \vec{\nabla}_f (X) |_{g(x)}$:

$$\begin{aligned} \vec{\nabla}_{\bar{\ell}}(\vec{\alpha}) &= \text{Using the MV Chain Rule with } \forall j,\ \vec{g}_j(\vec{x}) = \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{x} \right), f_j(\vec{y}) = \vec{y}^\top (\Delta t_j \Sigma)^{-1} \vec{y} \\ &- \frac{1}{2T} \sum_{j=0}^{T-1} \left(-\Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \right)^\top 2\Sigma^{-1} \Delta t_j^{-1} \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha} \right) \\ &= \frac{\Sigma^{-1}}{2T} \sum_{j=0}^{T-1} \begin{bmatrix} I & C_{t_j} \end{bmatrix}^\top \Sigma^{-1} \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha} \right) = \vec{0} \Rightarrow \text{left multiply by } T\Sigma \end{aligned}$$

$$\sum_{j=0}^{T-1} \begin{bmatrix} I \\ C_{t_j} \end{bmatrix} \Delta S_{t_j} = \sum_{j=0}^{T-1} \begin{bmatrix} I & C_{t_j} \end{bmatrix}^\top \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha} \Rightarrow$$

$$\sum_{j=0}^{T-1} \begin{bmatrix} I \\ C_{t_j} \end{bmatrix} \Delta S_{t_j} = \left(\sum_{j=0}^{T-1} \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix}^\top \begin{bmatrix} I & C_{t_j} \end{bmatrix} \right) \vec{\alpha} \Rightarrow$$

$$\hat{\alpha} = \left(\sum_{j=0}^{T-1} \Delta t_j \begin{bmatrix} I & C_j \\ C_j^\top & C_j^\top C_j \end{bmatrix} \right)^{-1} \left(\sum_{j=0}^{T-1} \begin{bmatrix} I \\ C_{t_j} \end{bmatrix} \Delta S_{t_j} \right)$$

Likewise, starting from 8 and noting $\frac{d\ \ln(\text{abs}(|A|))}{dA} = -\left(A^\top\right)^{-1}$, $\Sigma \succ 0 \to \ln\left(|\Sigma|\right) \in \mathbb{R}$, and $\frac{d\ \vec{x}^\top A^{-1} \vec{x}}{dA} = -\left(A^\top\right)^{-1} \vec{x}\vec{x}^\top \left(A^\top\right)^{-1}$:

$$\nabla_{\bar{\ell}}(\Sigma) = \frac{-\left(\Sigma^\top\right)^{-1}}{2} - \frac{1}{2T} \sum_{j=0}^{T-1} \frac{-\left(\Sigma^\top\right)^{-1} \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha}\right) \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha}\right)^\top \left(\Sigma^\top\right)^{-1}}{\Delta t_j}$$

$$= \frac{-\left(\Sigma^\top\right)^{-1}}{2} + \frac{\left(\Sigma^\top\right)^{-1}}{2T} \left(\sum_{j=0}^{T-1} \frac{\left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha}\right) \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha}\right)^\top}{\Delta t_j} \right) \left(\Sigma^\top\right)^{-1} = 0 \Rightarrow$$

$$\left(\Sigma^\top\right)^{-1} = \frac{\left(\Sigma^\top\right)^{-1}}{T} \left(\sum_{j=0}^{T-1} \frac{\left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha}\right) \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha}\right)^\top}{\Delta t_j} \right) \left(\Sigma^\top\right)^{-1} = 0 \Rightarrow$$

Left multiply by $\Sigma^\top$, then right multiply by $\Sigma^\top$

$$\Sigma^\top = \frac{1}{T} \sum_{j=0}^{T-1} \frac{\left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha}\right) \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha}\right)^\top}{\Delta t_j} \therefore$$

Under the assumption $\Sigma \succ 0$, it is desirable for $\hat{\Sigma}$ to also be PD $\Rightarrow \hat{\Sigma} = \hat{\Sigma}^\top$

$$\hat{\Sigma} = \frac{1}{T} \sum_{j=0}^{T-1} \frac{\left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha}\right) \left(\Delta \vec{S}_{t_j} - \Delta t_j \begin{bmatrix} I & C_{t_j} \end{bmatrix} \vec{\alpha}\right)^\top}{\Delta t_j}$$

4.2.2 MLE for the Geometric Growth Model

This subsection will differ from the previous discussion of the MLE parameters for the Arithmetic Growth Model in a major respect: the author was unable to find a closed form solution to the MLE for the **Dependent Geometric Growth Model** or **Dependent + Covariate Geometric Growth Model**. Discussion of this challenge and a practical resolution will be discussed at the end of this subsection.

First, (6) will be solved to find an expression for the likelihood of students' growth timeseries for the i^{th} cohort's dataset $\mathcal{D}_i$. The author would like to reiterate that the probability density of (6) and not (3.2.2) is modeled to fully account for Itô's Rules. The likelihood of $S_{\vec{t+\Delta t}}$ given $t, \vec{S}_t, C_t, \Delta_t$ is then:

Isolating the random portion of (6) applied at the j^{th} time interval $[t_j, t_{j+1}]$ of the i^{th} subdataset, the probability density of $\vec{S_{t_{j+1}}}$ given $t_{j+1}, t_j, \vec{S_{t_j}}, C_{t_j}$ can be modeled:

$$\vec{S_{t+\Delta t}} \sim \vec{S_t} \cdot \exp\left\{\begin{bmatrix} I & C_t \end{bmatrix} \begin{bmatrix} \vec{\mu} - \frac{1}{2}\mathrm{diag}\left(\Sigma\Sigma^T\right) \\ \theta \end{bmatrix} \Delta t + \Sigma\mathcal{N}(\vec{0}, \Delta t I)\right\} \Rightarrow$$

$$\ln\left(\vec{S_{t+\Delta t}}\right) - \ln\left(\vec{S_t}\right) \sim \begin{bmatrix} I & C_t \end{bmatrix} \begin{bmatrix} \vec{\mu} - \frac{1}{2}\mathrm{diag}\left(\Sigma\Sigma^T\right) \\ \theta \end{bmatrix} \Delta t + \Sigma\mathcal{N}(\vec{0}, \Delta t I) \Rightarrow$$

Let $\Delta\vec{X}_t := \ln\left(\vec{S_{t+\Delta t}}\right) - \ln\left(\vec{S_t}\right)$

$$\Delta\vec{X}_t - \begin{bmatrix} I & C_t \end{bmatrix} \begin{bmatrix} \vec{\mu} - \frac{1}{2}\mathrm{diag}\left(\Sigma\Sigma^T\right) \\ \theta \end{bmatrix} \Delta t \sim \mathcal{N}(\vec{0}, \Delta t\Sigma) \Rightarrow$$

$$f(\mathcal{D}_i j \mid \vec{\mu}, \Sigma, \vec{\theta}) = (2\pi)^{\frac{-N_{stu}}{2}} \cdot |\Delta t_j \Sigma|^{\frac{-1}{2}} .$$

$$= (2\pi)^{\frac{-N_{stu}}{2}} \cdot |\Delta t_j \Sigma|^{\frac{-1}{2}}$$

$$\cdot \exp\left\{\frac{-1}{2\Delta t_j}\left(\Delta\vec{X}_t - \begin{bmatrix} I & C_t \end{bmatrix}\begin{bmatrix} \vec{\mu} - \frac{\mathrm{diag}\left(\Sigma\Sigma^T\right)}{2} \\ \theta \end{bmatrix}\Delta t\right)^{\top} \Sigma^{-1} \left(\Delta\vec{X}_t - \begin{bmatrix} I & C_t \end{bmatrix}\begin{bmatrix} \vec{\mu} - \frac{\mathrm{diag}\left(\Sigma\Sigma^T\right)}{2} \\ \theta \end{bmatrix}\Delta t\right)\right\} \therefore$$

The log –likelihood of the j^{th} interval is

$$\ell(\mathcal{D}_{ij} \mid \vec{\mu}, \Sigma, \vec{\theta}) = -\frac{N_{stu}}{2}\ln(2\pi) - \frac{\ln(\Delta t_j) + \ln(|\Sigma|)}{2}$$

$$- \frac{\left(\Delta\vec{X}_t - \begin{bmatrix} I & C_t \end{bmatrix}\begin{bmatrix} \vec{\mu} - \frac{\mathrm{diag}\left(\Sigma\Sigma^T\right)}{2} \\ \theta \end{bmatrix}\Delta t\right)^{\top} \Sigma^{-1} \left(\Delta\vec{X}_t - \begin{bmatrix} I & C_t \end{bmatrix}\begin{bmatrix} \vec{\mu} - \frac{\mathrm{diag}\left(\Sigma\Sigma^T\right)}{2} \\ \theta \end{bmatrix}\Delta t\right)}{2\Delta t_j} \Rightarrow$$

Because BM is the only modeled source of randomness in the evolution of $\vec{S}_t$ and BM has an independent increments property of BM, the likelihood of the timeseries is the product of the likelihood of its intervals. Thus, the average log –likelihood of the i^{th} subdataset is

$$\bar{\ell}(\mathcal{D}_{ij} \mid \vec{\mu}, \Sigma, \vec{\theta}) = -\frac{N_{stu}}{2}\ln(2\pi) - \frac{\sum_{j=0}^{T}\ln(\Delta t_j)}{2T} - \frac{\ln(|\Sigma|)}{2}$$

$$- \frac{1}{2T}\sum_{j=0}^{T-1} \frac{\left(\Delta\vec{X}_t - \begin{bmatrix} I & C_t \end{bmatrix}\begin{bmatrix} \vec{\mu} - \frac{\mathrm{diag}\left(\Sigma\Sigma^T\right)}{2} \\ \theta \end{bmatrix}\Delta t\right)^{\top} \Sigma^{-1} \left(\Delta\vec{X}_t - \begin{bmatrix} I & C_t \end{bmatrix}\begin{bmatrix} \vec{\mu} - \frac{\mathrm{diag}\left(\Sigma\Sigma^T\right)}{2} \\ \theta \end{bmatrix}\Delta t\right)}{\Delta t_j} \tag{9}$$

For the **Independent Geometric Growth Model**, $\vec{\mu}$ and Σ are scalars (redenoted as μ, σ), and the mean log likelihood of $\mathcal{D}_i$ is

$$\bar{\ell}(\mathcal{D}_{ij} \mid \mu, \sigma) = -\frac{N_{stu}}{2}\ln(2\pi) - \frac{\sum_{j=0}^{T}\ln(\Delta t_j)}{2T} - \frac{\ln(\sigma)}{2} - \frac{1}{2T}\sum_{j=0}^{T-1}\frac{\left(\Delta\vec{X}_t - \left(\mu - \frac{\sigma^2}{2}\right)\Delta t\right)}{\Delta t_j \sigma}$$

Maximizing the likelihood by setting the partial derivative of the parameters to 0, gives the system $\begin{bmatrix} \vec{\nabla}_{\bar{\ell}}(\mu) \\ \vec{\nabla}_{\bar{\ell}}(\sigma) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, which can be solved with careful application of elementary calculus and algebra:

$$\hat{\mu} = \frac{\delta X}{\delta t} + \frac{\hat{\sigma}^2}{2} \tag{10}$$

$$\hat{\sigma}^2 = 2\frac{\sqrt{\delta t(\mu^2\delta t - 2\mu\delta X + \sum_n \Delta X_n^2 \Delta t_n^{-1}) + N^2} - N}{\delta t} \tag{11}$$

where $\delta X := \ln\left(S_{t_f}\right) - \ln\left(S_{t_0}\right)$ and $\delta t := t_f - t_0$.

For the **Dependent Geometric Growth Model** and **Dependent + Covariate Geometric Growth Model**, however, the term $\mathrm{diag}\left(\Sigma\Sigma^T\right)$ makes manual differentiation difficult. Note that

$$\mathrm{diag}\left(\Sigma\Sigma^T\right) = \sum_{i=1}^{N_{stu}} \left(\hat{e}_i{}^\top \Sigma\Sigma\hat{e}_i\right)\hat{e}_i$$

$\because \Sigma \succ 0 \to \Sigma = \Sigma^\top$

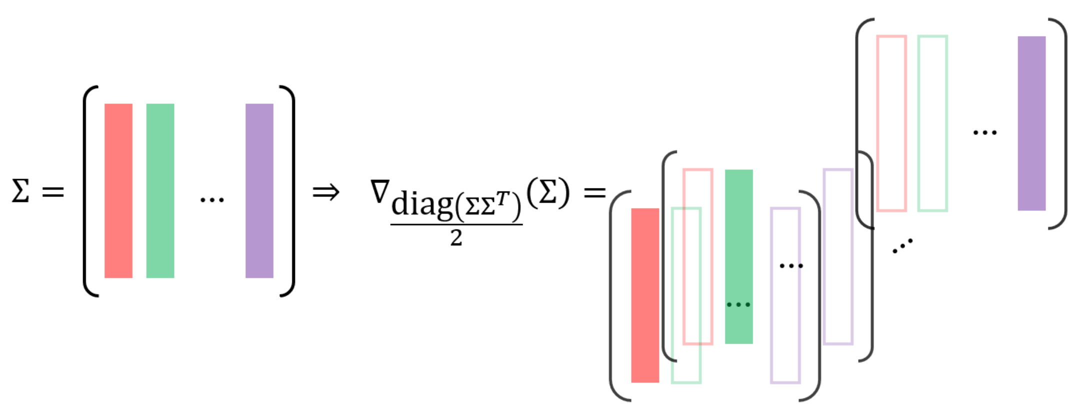
$$= \sum_{i=1}^{N_{stu}} \left(\hat{e}_i{}^\top \Sigma^\top \Sigma\hat{e}_i\right)\hat{e}_i \Rightarrow$$

$$\vec{\nabla}_{\mathrm{diag}(\Sigma\Sigma^T)}(\Sigma) = \sum_{i=1}^{N_{stu}} \vec{\nabla}_{\left(\hat{e}_i{}^\top \Sigma^\top\Sigma\hat{e}_i\right)\hat{e}_i}(\Sigma)$$

MV Chain Rule for $f(g(A))$ with $g(A) = (A\vec{v})^\top (A\vec{v}), f(c) = c\vec{v}$ and $\vec{v} = \hat{e}_i \wedge A = \Sigma$

$$= \sum_{i=1}^{N_{stu}} \left(2\Sigma\hat{e}_i\hat{e}_i{}^\top\right)^\top \otimes \hat{e}_i \tag{12}$$

In particular, the tensor product $\otimes$ adds complexity, but is needed since the gradient of vector valued fucntion $\mathrm{diag}\left(\Sigma\Sigma^T\right) \in \mathbb{R}^{N_{stu}}$ with respect to matrix $\Sigma \in \mathbb{R}^{N_{stu}\times N_{stu}}$ will give a tensor in $\mathbb{R}^{N_{stu}\times N_{stu}\times N_{stu}}$. There is a lot of inherent simplicity in (12). For example, $\Sigma\hat{e}_i\hat{e}_i{}^\top$ is the i^{th} column of Σ mapped onto the i^{th} column of a $N_{stu} \times N_{stu}$ 0 matrix. The $\otimes\hat{e}_i$ stacks this sparse matrix into the i^{th} index of an otherwise 0 tensor. After taking the sum across i in $\sum_{i=0}^{N_{stu}}$, $\vec{\nabla}_{\mathrm{diag}(\Sigma\Sigma^T)}(\Sigma)$ can be visualized as:



The author supposes that these properties of $\vec{\nabla}_{\text{diag}(\Sigma\Sigma^T)}(\Sigma)$ can be better understood, and a closed form solution for the parameters of **Dependent Geometric Growth Model**s can be found. Such a derivation is an open question for the mathematical community and the author will continuously publish more insight as his mathematical maturity grows.

For completion of this project, however, computational methods are required to find parameters for the **Dependent Geometric Growth Model** and **Dependent + Covariate Geometric Growth Model**.

4.2.3 Iterative Methods for Learning the Parameters of Dependent Geometric Growth Models

Since no closed form solution to the maximum of (9) could be found in the previous method, gradient descent is used to estimate $\vec{\mu}, \Sigma, \vec{\theta}$ (when applicable) that maximize (9) on any $\mathcal{D}_{train,i}$.

As Nesterov 2004 notes in *Introductory Lectures on Convex Optimization* (see eqn 2.1.17 on page 66), the error of the k^{th} epoch of gradient descent is proportional to the square of the norm of difference between the initial parameters and optimal parameters (with the additional constraint that the loss function is smooth in the path between these parameters). Thus, careful initialization of $\vec{\mu_0}, \Sigma_0, \vec{\theta_0}$ can reduce the number of iterations needed to converge to the final estimates of $\hat{\mu}, \hat{\Sigma}, \hat{\theta}$.

Note that the only difference between the average log –likelihood functions of the **Dependent + Covariate Arithmetic Growth Model** (8) and **Dependent + Covariate Geometric Growth Model** (9) are the $d\vec{X}_{t_j} := \ln\left(\vec{S_{t_{j+1}}}\right) - \ln\left(\vec{S_{t_j}}\right)$ and $\frac{\text{diag}(\Sigma\Sigma^T)}{2}$ terms in the latter (in place of $d\vec{S}_{t_j} := \vec{S_{t_{j+1}}} - \vec{S_{t_j}}$ and $\varnothing$ in the former). Thus, one could get parameters $\hat{\mu}_{\text{Arith}}, \hat{\Sigma}_{\text{Arith}}, \hat{\theta}_{\text{Arith}}$ by regressing $\ln\left(\vec{S_t}\right)$ on $\mathcal{D}_{train}$ and seed gradient descent for the **Dependent + Covariate Arithmetic Growth Model** as:

$$
\begin{aligned}
\vec{\mu_0} &= \hat{\mu}_{\text{Arith}} - \frac{\hat{\Sigma}_{\text{Arith}}\hat{\Sigma}^{\top}_{\text{Arith}}}{2} \\
\Sigma_0 &= \hat{\Sigma}_{\text{Arith}} \\
\vec{\theta_0} &= \hat{\theta}_{\text{Arith}}
\end{aligned}
\tag{13}
$$

and run gradient descent until an evaluation metric (e.g., mean absolute error) on $\mathcal{D}_{val}$ diverges from that metric on $\mathcal{D}_{train}$. This process will be referred to as the **log–Arithmetic Initialization Trick**

The learning rates for $\vec{\mu_k}$ and Σ_k are initialized at 1e-3 and the learning rate for $\vec{\theta_k}$ was initialized to 1e-4. This difference is because $\vec{\theta_k}$ is learned across all 14 subdatasets $\mathcal{D}_{train_i}$ while $\vec{\mu_k}$ and Σ_k are reinitialized every time a new $\mathcal{D}_{train_i}$ is loaded. During the reinitialization of $\vec{\mu_k}$ and Σ_k, the **log–Arithmetic Initialization Trick**is used with the caveat that $\vec{\theta_k}$ is facored into calculations of $\hat{\mu}_{\text{Arith}}$ and $\hat{\Sigma}_{\text{Arith}}$ instead of a simultaneous calculation of $\hat{\theta}_{\text{Arith}}$.

Results

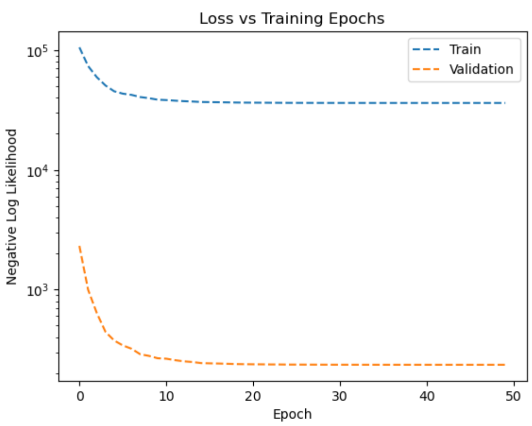
The code for this thesis is available on Github. Note that the code and examples on this Github repository use financial terms and data because the author does not have permission to publicly post students data.

Please contact Mike Preiner at The Math Agency for inquiries about accessing the specific datasets used in this report.

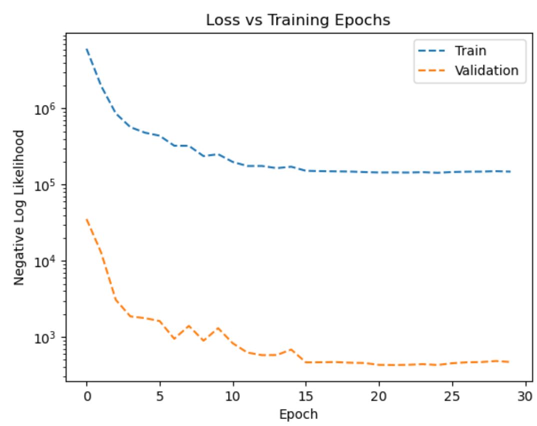
5.1 Empirically Understanding the log–Arithmetic Initialization Trick

Before presenting the analytical results, it is essential to check that the iterative schemes for the **Dependent Geometric Growth Models** do converge in a consistent, well-behaved manner.

In practice, the chosen iteration method was mini-batch gradient descent seeded with the **log–Arithmetic Initialization Trick**(13). Mini-batch gradient descent was preferred over gradient descent because in practice, it is helpful to analyze student's growth in the most recent two months. **Mini-batch gradient descent with batches of training data split along the time axis allows the model to function this way during training time as well.** Additionally learning rate decay schedule helped prevent the loss from exploding on any dataset (the schedule chosen was to decrease the learning rate after every five epochs over the training dataset such that the learning rate decays by a factor of ten every ten epochs).



(a) Representative Loss Curve for the **Dependent Geometric Growth Model**



(b) Loss Curve for **Dependent + Covariate Geometric Growth Model**

Additionally, the observation that the training curve has flattened suggests that after 50 epochs, the resulting model is the MLE **Dependent Geometric Growth Model** for this subdataset's training split.

The training of the **Dependent Geometric Growth Model** parameters results in the following models:

As time increases, the range of future predictions become more concentrated. The bias, however, does not change with additional training, suggesting that this model might not be a good way to forecast student progress.

5.2 Model Comparison

To empirically answer the question of which model has the greatest utility for forecasting student progress, four metrics are considered:

a. mean absolute error (MAE) across forecasted growth signals

b. MAE at the final forecasted growth signal

c. root mean squared error (RMSE) across forecasted growth signals

d. RMSE at the final forecasted growth signal

Note that MAE is a unit-less measure, but the unit of RMSEs in the below table is "grades of achievement."

The following table presents 99% Confidence Intervals for each model (N = 10 simulations), metric pairing. Each point estimate used in these metrics' calculation involves all random aspects of this study (stochastic imputation (see 4.1), random noise added to ensure all initial covariance matrices $\hat{\Sigma}_0$ are almost surely invertible, etc.)

Table 1: MAE on Data Splits for Trained Models

Growth Model	Train	Val	Test
Indep. Arith.	[0.0557, 0.0580]	[0.0799, 0.0817]	[0.0551, 0.0576]
Dep. Arith.	[0.0317, 0.0337]*	[0.0695, 0.0715] †	[0.0424, 0.0451] †
Dep.+Cov. Arith.	[0.0355, 0.0381]	[0.0687, 0.0710] †	[0.0403, 0.0432] †
Indep. Geom.	[0.0683, 0.0697]	[0.2984, 0.3204]	[0.0556, 0.0594]
Dep. Geom.	[0.0683, 0.0695]	[0.1381, 0.1469]	[0.0466, 0.0497]
Dep.+Cov. Geom.	[0.2233, 0.2410]	[0.1518, 0.1633]	[0.1877, 0.1995]

Table 2: Last-Timestamp MAE on Data Splits for Trained Models

Growth Model	Train	Val	Test
Indep. Arith.	[0.0312, 0.0322]	[0.1026, 0.1058]	[0.0816, 0.0886]
Dep. Arith.	[0.0199, 0.0218]*	[0.0885, 0.0921] †	[0.0671, 0.0697] †
Dep.+Cov. Arith.	[0.0247, 0.0255]	[0.0884, 0.0911] †	[0.0637, 0.068] †
Indep. Geom.	[0.1016, 0.1171]	[0.6464, 0.6914]	[0.0894, 0.0963]
Dep. Geom.	[0.0667, 0.0716]	[0.2421, 0.2581]	[0.0781, 0.0832]
Dep.+Cov. Geom.	[0.2078, 0.2286]	[0.2427, 0.2624]	[0.3386, 0.3574]

Table 3: RMSE on Data Splits for Trained Models

Growth Model	Train	Val	Test
Indep. Arith.	[0.0449, 0.0466]	[0.1367, 0.1393]	[0.1052, 0.1098]
Dep. Arith.	[0.0258, 0.0272]*	[0.1157, 0.1187] †	[0.0816, 0.0904] †
Dep.+Cov. Arith.	[0.0283, 0.0296]	[0.1158, 0.1194] †	[0.0774, 0.0853] †
Indep. Geom.	[0.0610, 0.0633]	[0.5164, 0.5559]	[0.1057, 0.1125]
Dep. Geom.	[0.0649, 0.0686]	[0.2379, 0.2534]	[0.0916, 0.0994]
Dep.+Cov. Geom.	[0.1481, 0.1645]	[0.2484, 0.2657]	[0.3538, 0.3835]

Table 4: Last-Timestamp RMSE on Data Splits for Trained Models

Growth Model	Train	Val	Test
Indep. Arith.	[0.0511, 0.0527]	[0.1754, 0.1818]	[0.1491, 0.166]
Dep. Arith.	[0.0471, 0.0533]	[0.1464, 0.1508] †	[0.1272, 0.1333]
Dep.+Cov. Arith.	[0.0272, 0.0292]*	[0.1454, 0.1498] †	[0.1156, 0.1249]*
Indep. Geom.	[0.1549, 0.1879]	[1.1363, 1.2617]	[0.1656, 0.1808]
Dep. Geom.	[0.1043, 0.1115]	[0.4230, 0.4630]	[0.1514, 0.1666]
Dep.+Cov. Geom.	[0.2588, 0.2846]	[0.3668, 0.3951]	[0.5587, 0.5988]

Note: * consensus best model for this metric †model's CI for this metric captures the lowest upper bound

5.3 Covariate Coefficients

Since the **Dependent Arithmetic Growth Models** outperformed all other models on all test metrics, the coefficients for the covariates in the **Dependent + Covariate Arithmetic Growth Model** might be of interest. The following table includes 99% Confidence Intervals for $\hat{\theta}$ after $N = 50$ simulations.

Table 5: Covariate Coefficients under the **Dependent + Covariate Arithmetic Growth Model**

Covariate i	CI for $\hat{\theta}_i$
Practice Duration (mins)	$[7.0674e-05,\ 2.5800e-04]$
Athome Practice Duration (mins)	$[-9.0245e-06,\ 2.3585e-04]$
Num. Activities Practiced	$[-9.7949e-06,\ 2.2818e-04]$
Percent Activities Finished	$[-2.4026e-04,\ 2.4142e-04]$
Num. Restarts per Activity	$[-1.5964e-04,\ 1.9375e-04]$
Attendance Rate	$[-1.8741e-04,\ 2.0236e-04]$
Attendance Rate (past 2 months)	$[-2.1143e-04,\ 2.4217e-04]$
is in Grade 2	$[-2.0888e-04,\ 2.1791e-04]$
is in Grade 3	$[-1.8650e-04,\ 1.7673e-04]$
is in Grade 4	$[-1.9320e-04,\ 2.1373e-04]$
is in Grade 5	$[-2.1298e-04,\ 1.6481e-04]$

Interestingly, only the coefficient for *Practice Duration (mins)* was significant at the $\alpha = 1\%$ level. (This may also explain why the performance for the **Dependent Arithmetic Growth Model** and **Dependent + Covariate Arithmetic Growth Model** were so similar).

Discussions and Conclusions

6.1 Final Model

Because the **Dependent Arithmetic Growth Model** and **Dependent + Covariate Arithmetic Growth Model** outperformed all other models on all metrics (see Tables 1-4 in 5.2) and the only significant covariate was *Practice Duration (mins)*, this author proposes the following model for student growth.

$$\vec{S_{t+\Delta t}} = \vec{S}_t + \left(\vec{\mu} + 1.6434\text{e-04} \cdot \vec{\text{practice}}\ \text{mins}\right)\Delta t + \mathcal{N}\left(0, \Delta\Sigma\right) \tag{14}$$

The better fit of **Arithmetic Growth Models** suggests that students are learning some stochastic percentage of a concept daily, regardless of the size of thier existing knowledge base. There is no evidence in this study to suggest that children inherently learn math this manner. Perhaps pedagogical practices can help students use a greater portion of their math knowledge, connect subdisciplines like algebra and geometry more intuitively, and help students learn at measurably exponential rates.

6.2 Implications and Applications for Classroom Management

The parameters $\vec{\mu}$ and Σ from this model can be used to assist teachers in everyday classroom tasks. For example, for each student i, their growth rate adjusted for practice time is μ_i and the volatility in their growth is $\sigma_i := \|\Sigma_{i,:}\|$.

Running this model on a Math Agency cohort that the author is familiar with:

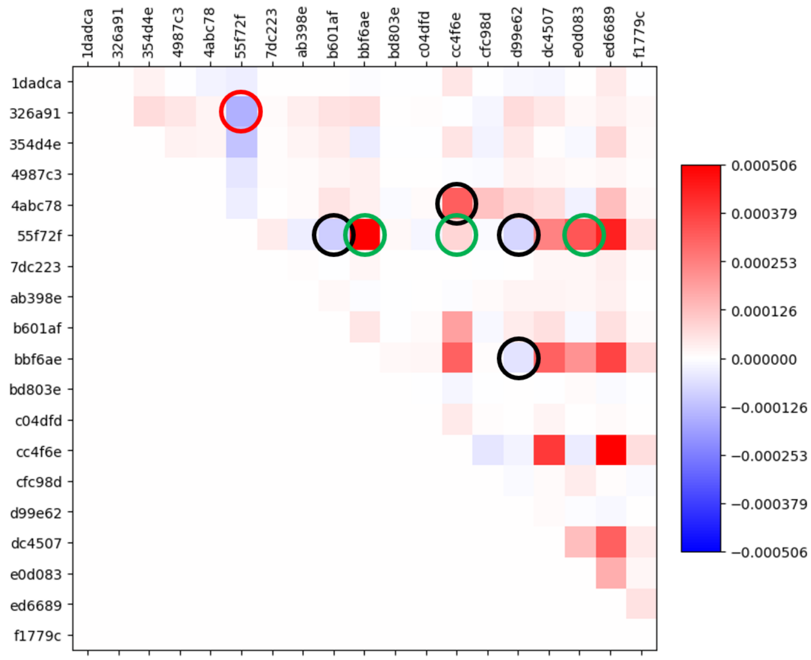
Table 6: Student Growth Parameters

Student i's ID	$\hat{\mu}_i$ under (14)	$\hat{\sigma}_i$ under (14)	$\hat{\mu}_i$ under **Dep. AGM** (3)	$\hat{\sigma}_i$ under (3)
1dadca	4.97e-04	1.01e-04	2.58e-03	1.00e-04
326a91	-1.22e-03	2.52e-04	2.63e-03	2.63e-04
354d4e	3.47e-03	2.44e-04	7.84e-03	3.13e-04
4987c3	-1.94e-03^{5}	9.70e-05	1.13e-03^{5}	1.19e-04
4abc78	4.75e-03	6.43e-04	8.35e-03	6.87e-04
55f72f	7.49e-03	1.75e-03^{1}	9.79e-03	1.83e-03^{1}
7dc223	-3.47e-03^{3}	6.10e-05	6.08e-04^{3}	6.30e-05
ab398e	-3.17e-03^{4}	7.70e-05	2.64e-03	1.26e-04
b601af	-3.29e-04	2.60e-04	2.07e-03	3.03e-04
bbf6ae	5.13e-03	1.13e-03^{3}	1.15e-02	1.35e-03^{3}
bd803e	-1.17e-03	3.20e-05	5.63e-04^{2}	4.50e-05
c04dfd	-3.60e-03^{2}	5.70e-05	5.08e-04^{1}	6.40e-05
cc4f6e	7.66e-03	1.71e-03^{2}	9.89e-03	1.80e-03^{2}
cfc98d	1.27e-03	1.77e-04	3.16e-03	2.05e-04
d99e62	1.96e-03	2.07e-04	3.10e-03	2.05e-04
dc4507	5.62e-03	7.62e-04^{5}	8.84e-03	8.67e-04^{5}
e0d083	6.34e-04	4.83e-04	4.00e-03	5.14e-04
ed6689	6.06e-03	1.05e-03^{4}	1.19e-02	1.13e-03^{4}
f1779c	-3.68e-03^{1}	1.33e-04	6.89e-04^{4}	1.42e-04

Note: • the units of μ_i and σ_i are grades/day. • A number k next to metric means that a student is the k^{th} h
(i.e., a teacher wants to help the students with lowest growth rates μ_i and highest volatilities σ_i)

The addition of the additional covariate does not change the ordering of student's volatilities, but does make it easier to flag students who are putting in extra effort (i.e., are spending addtional time practicing at home or are extra focused during Math Agency sessions) to learn but not fully benefitting from it. For example, the student with ID ab398e has a high growth rate when directly measured but (from experience teaching this student) would learn more per unit time spent practicing with faster calculation abilities. If educators are given reminders to check on such students and teach them new strategies, then these students will feel motivated by growing at a rate commensurate with their efforts.

The parameter Σ is can also be used when assigning table grouping to students. The matrix $\hat{\Sigma}$ for the cohort discussed above is:



Off-Diagonal Entries of $\hat{\Sigma}$

The red circle is statistical evidence that students with IDs 326a91 and 55f2f do not work well together. For those students, a statistical analysis was not necessary to reach taht conclusion. The green circles are evidence that the student with ID 55f2f (who causes class wide disruptions more than peers but has also become a quite skilled mathemetician over the school year) positively reinforces his friends' growth. When assigning seating, teachers could choose to group these students together to see if this trend will continue (or if it was because of their separation), potentially helping all four friends grow.

It is, however, possible to over analyze the entries $\hat{\Sigma}_i j$. For example, the student pairs circled in black have strong positive or negative covariance. These students, however, sit in separate classrooms during Math Agency sessions.

6.3 Limitations and Uncertainty Analysis

As with any data-driven analysis, the results from this study should not be generalized to students or programs different from The Math Agency's. The methods from this study, however, can be applied to any domain where there are (potentially) correlated processes that evolve over time.

The addition of covariate data to SDE based models did not have a very large impact on this study. It is possible that unobserved covariates play an important role in student's growth timeseries. For example, earlier in this section, there was discussion about a student who would benefit from practicing fact fluency (small sums, times tables) in a timed environment. The Math Agency recently started piloting the xtramath platform to collect such data, so within the next couple years, there should be enough data to incorporate this information in analyzes similar to the present one. The possibilities and descriptive power of **Dependent + Covariate Arithmetic Growth Models** will grow as the number of covariates collected in a high-quality manner grows.

There are also inherent limitations in the SDE models chosen for this project. One main assumption was that parameters $\vec{\mu}$ and Σ are latent to students and are invariant to time or other factors. A class of SDE-based models called diffusion models learns $\hat{\mu}$ and $\hat{\Sigma}$ using a subset of the sigma field $\sigma(S_t)$ to predict $S_{t+\Delta t}$ (see, e.g., Lim et al. 2023 and Lugmayr et al. 2022).

Acknowledgements

The paper's author is extremely grateful to Dr. Larry Pierce of the University of Washington Department of Mathematics for his mentorship throughout the development for this paper.

The author would like to thank Mike Preiner at The Math Agency for remarks that led to the development of the Arithmetic Growth Models' SDEs.

References

Black, Fischer and Myron Scholes (1973). "The Pricing of Options and Corporate Liabilities". In: *Journal of Political Economy* 81.3, pp. 637–654. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/1831029 (visited on 05/19/2025).

Goldhaber, Dan and Maia Goodman Young (2024). "Course grades as a signal of student achievement: Evidence of grade inflation before and after COVID-19". In: *Journal of Policy Analysis and Management* 43.4. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pam.22618, pp. 1270–1282. ISSN: 1520-6688. DOI: 10.1002/pam.22618. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pam.22618 (visited on 04/28/2025).

Goldhaber, Dan D., Dominic J. Brewer, and Deborah J. Anderson (Dec. 1999). "A Three-way Error Components Analysis of Educational Productivity". In: *Education Economics* 7.3, pp. 199–208. ISSN: 0964-5292. DOI: 10.1080/09645299900000018. URL: https://doi.org/10.1080/09645299900000018 (visited on 04/28/2025).

Itô, Kiyosi (1944). "Stochastic integral". In: *Proceedings of the Imperial Academy* 20.8, pp. 519–524. DOI: 10.3792/pia/1195572786. URL: https://doi.org/10.3792/pia/1195572786.

— (1951). "On a Formula Concerning Stochastic Differentials". In: *Nagoya Mathematical Journal* 3, pp. 55–65. DOI: 10.1017/S0027763000012216.

Lim, Sungbin et al. (2023). "Score-based generative modeling through stochastic evolution equations in hilbert spaces". In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. NIPS '23. New Orleans, LA, USA: Curran Associates Inc.

Lugmayr, Andreas et al. (June 2022). "RePaint: Inpainting Using Denoising Diffusion Probabilistic Models". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11461–11471.

Math Agency, The (2024). *2023-24 Annual Social Purpose Report*. URL: https://static1.squarespace.com/static/64db9af4e7e8fa3da5ee67cc/t/67ca4e0bc3abe909bbb1efb6/1741311501095/2023-24+Annual+Social+Purpose+Report.pdf.

Merton, Robert C. (1973). "Theory of Rational Option Pricing". In: *The Bell Journal of Economics and Management Science* 4.1. Publisher: [Wiley, RAND Corporation], pp. 141–183. ISSN: 0005-8556. DOI: 10.2307/3003143. URL: https://www.jstor.org/stable/3003143 (visited on 04/28/2025).

Nesterov, Yurii (2004). *Introductory Lectures on Convex Optimization*. eng. Applied Optimization. New York: Springer. ISBN: 978–1–4419–8853–9.

Preiner, Mike (Aug. 2022). *Can we measure the efficiency of math practice?* URL: https://mpreiner.medium.com/can-we-measure-the-efficiency-of-math-practice-c4173d59d84e (visited on 04/28/2025).

Shreve, Steven E. (2004). *Stochastic calculus for finance*. eng. Springer finance. Textbook. New York: Springer. ISBN: 0387401008.

Zhang, Yanwei and Liru Zhang (Apr. 2002). "Modeling School and District Effects in the Math Achievement of Delaware Students Measured by DSTP: A Preliminary Application of Hierarchical Linear Modeling in Accountability Study". In: URL: https://www.academia.edu/69029769/Modeling_School_and_District_Effects_in_the_Math_Achievement_of_Delaware_Students_Measured_by_DSTP_A_Preliminary_Application_of_Hierarchical_Linear_Modeling_in_Accountability_Study (visited on 05/16/2025).