

# Investigating the Impact of Slipping Parameters on Additive Factors Model Parameter Estimates

Christopher J. MacLellan  
Carnegie Mellon University  
5000 Forbes Avenue  
Pittsburgh, PA 15213  
cmaclell@cs.cmu.edu

## ABSTRACT

The Additive Factors Model (AFM), a widely used model of student learning, estimates students' prior knowledge, the difficulty of tutored skills, and the rates at which these skills are learned. In contrast to Bayesian Knowledge Tracing (BKT), another widely used model of student learning, AFM does not have parameters for the slipping rates of learned skills; i.e., it does not explicitly model situations where students know a skill, but still apply it incorrectly. Thus, AFM assumes that as students get more practice their probability of correctly applying a skill converges to 100%, whereas BKT allows convergence to lower probabilities. This restriction constrains the range of values that AFM parameters can take. In particular, when the asymptotic performance of a skill is less than 100%, AFM will estimate the learning rate to be lower than if slipping was taken into account. To investigate this phenomenon, I will create a LearnSphere workflow component that implements AFM and a variant of AFM with explicit slipping parameters (AFM+S). Using this component, I analyze multiple DataShop datasets to determine (1) whether the model with slipping parameters better fits the data and (2) how the addition of slipping parameters impacts the parameter estimates returned by AFM. I show that, in general, AFM+S better fits the data than the AFM. Additionally, I show that AFM+S estimates higher skill intercepts and learning rates than AFM, whereas AFM estimates higher student intercepts than AFM+S.

## Keywords

Cognitive Modeling, Statistical Models of Learning, Additive Factors Model, Knowledge Tracing.

## 1. INTRODUCTION

The Additive Factors Model [1], or AFM, is a statistical model of student learning that can be fit to educational data in order to estimate students' prior knowledge, the difficulty of tutored skills, and the rates at which these skills are learned. Unlike Bayesian Knowledge Tracing [2], an alternative statistical model of student learning, AFM does not have explicit parameters to model the rate at which students incorrectly apply learned skills (i.e., slipping parameters).

This lack of slipping parameters has an impact on both the model fit and the parameter estimates. If slipping is occurring, then model fits should improve by taking these parameters into account. Further, in situations where slipping is occurring, AFM will underestimate learning rates so that it can fit the higher error rates in the tail of the learning curve [3]. There is some evidence that the learning rates estimated by BKT, an approach that takes slipping into account, tend to be higher than those estimated by

AFM [4]. However, a thorough investigation of how slipping rates impact learning estimates has not been done.

In order to investigate the impact of slipping parameters on AFM's model fit and parameter estimates, I created a LearnSphere workflow component that implements both AFM and the extension of AFM that includes slipping parameters. I refer to this extension as AFM+S [3]. Using this component I fit both AFM and AFM+S models to five datasets from DataShop. I analyzed the output to determine which model best fits the data, whether slipping was occurring in the datasets, and to compare the parameter estimates of the two models to determine how the slipping parameters affect the learning rate estimates.

Previous work has shown that AFM+S better fits the data better than AFM and BKT on five different datasets [3]. I replicated this analysis to show that the same results hold with the new workflow component. Further, in my analysis I took additional precautions to prevent Type I errors (i.e., identifying a significant difference when none exists). As a preliminary test that learning rates estimated by the AFM+S model will be higher than the learning rates estimated by the AFM model, I fit both models to the Geometry Area 1996-1997 dataset accessed via DataShop [5] and compared their learning rate estimates. I found that the mean learning rate for the AFM model was 0.18 logits, whereas the mean learning rate for the AFM+S model was 0.42, a significant difference ( $V=0$ ,  $p < 0.01$  via a paired Wilcoxon signed-rank test). These preliminary results suggested that adding slipping parameters to the model causes the estimated learning rates to be higher. However, I wanted to analyze the other four datasets to identify whether this was a systematic trend. In this paper I will present the results of this analysis. In particular, I show AFM+S better fits the five datasets than AFM on unstratified and stratified cross validation and that the skill intercepts and slopes (i.e., learning rates) estimated by the AFM+S model are higher than those estimated by the AFM model. Further, I also show that the AFM model estimates the student intercepts to be higher than the AFM+S model.

In addition to exploring these ideas, this paper showcases the new LearnSphere workflow component. Researchers can use this component in situations where they want to use AFM, but where they suspect slipping is occurring. BKT is one possible alternative, but is not a panacea. For example, BKT does not support multiple skill labels per step, but AFM+S does. Further, there is evidence that AFM+S better fits many datasets than the traditional BKT [3]. A workflow component for AFM+S is a contribution to the ecosystem of learning analytic models that researchers might like to use.

## 2. WORKFLOW COMPONENT

### 2.1 Data Inputs

The AFM+S workflow component that I am created accepts the standard PSLC DataShop student-step rollup format. From these files the AFM+S model requires information about the student labels, the knowledge component labels, and the knowledge component opportunity counts. Depending on whether item cross-validation is to be performed, the model also needs the item labels.

### 2.2 Workflow Model

The code for the AFM+S workflow component is implemented in Python and is publicly available on GitHub: <https://github.com/cmacllell/pyAFM>. This code implements a standard Logistic Regression classifier that accepts box-constraints (so learning rates can be constrained to be positive) and L2 regularization parameters (so student intercepts can be pulled towards 0). It also implements Bounded Logistic Regression, so that slipping parameters can be taken into account. Using these classifiers, the code provides implementations of both AFM and AFM+S as described in prior work [3].

### 2.3 Workflow Outputs

The AFM+S workflow component has three possible outputs. First, it outputs metrics for assessing the fit of the model to data. In particular, it outputs unstratified, stratified, student, and item cross-validated root-mean-square error. Second, the model outputs predicted first-attempt performance for each student step, so that the resulting learning curve can be plotted and compared to alternative models. Finally, the model outputs student intercept parameter estimates, skill difficulty and learning rate parameter estimates, and skill slipping parameter estimates.

The model fit statistics and parameter estimate outputs take the form of tables or comma-separated value output files. The model predictions output takes the form of either a comma-separated value output file or learning curve plots. These learning curve plots are similar to those currently available on DataShop.

## 3. METHOD

In order to investigate the impact of slipping parameters on AFM skill slopes, I used the new workflow component to fit both the AFM and AFM+S models to five datasets downloaded from DataShop: Geometry [5], Equation Solving [6,7], Number Line Estimation [8], Writing 1 [9], and Writing 2 [10].

Before analyzing parameter differences, I assessed which model better fit the data using cross validation. For each model and dataset, I performed 5 runs of 2-fold stratified and unstratified cross validation and 1 run of 2-fold student and item cross validation (i.e., where students and items are divided across the folds). I then used a Paired Wilcoxon Signed-Rank test to compare the model fits across the datasets, runs, and folds. I did not conduct more runs or folds because there is evidence that doing so increases the risk of Type I error due to the correlation in model fits between folds that share training data [11]. For student and item cross validation, I conducted only 1 run of 2 fold cross validation because randomly splitting students and items between fold, while balancing the number of training points between folds, is non-random and repeated runs also increases the likelihood of Type I error.

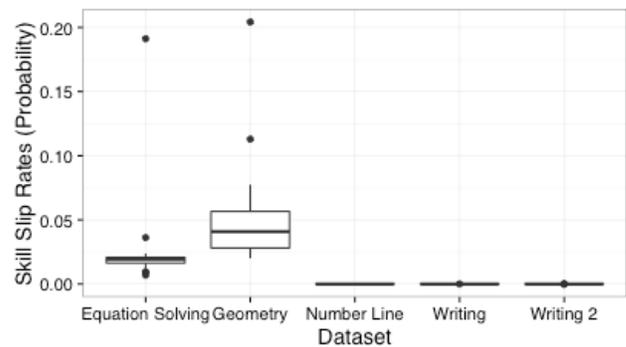


Figure 1. The slipping rates of skills across the five datasets.

After assessing overall model fits, I fit each model (AFM and AFM+S) to each of the datasets using all of the available data and recorded the parameter estimates from both models. I plotted the slipping parameter values to determine which datasets are most affected by the slipping parameters (Figure 1). In situations where there is little slipping, AFM+S should be identical to AFM. I then compared each of the parameter types (skill intercepts, skill slopes, and student intercepts) between models using a Paired Wilcoxon Signed-Rank test to determine if there were systematic differences in parameter estimates produced by the models across the five dataset.

## 4. RESULTS

Overall the AFM+S model better fits the data across the five datasets and four cross-validation types (unstratified, stratified, student, and item), via a Wilcoxon Signed-Rank Test paired by cross-validation type, dataset, run, and fold ( $V=1350.5$ ,  $p < 0.01$ ). When dividing the data by cross-validation type, AFM+S better fits the data across the five datasets for unstratified ( $V=213$ ,  $p < 0.01$ ) and stratified ( $V=222$ ,  $p < 0.01$ ), but not student ( $V=8$ ,  $p=1$ ) and item ( $V=26$ ,  $p > 0.7$ ) cross validation. When dividing the data by dataset, AFM+S better fits the data on Geometry ( $V=212$ ,  $p < 0.01$ ) and Equation Solving ( $V=181$ ,  $p < 0.02$ ), but not Number Line ( $V=3$ ,  $p = 1$ ), Writing 1 ( $V=2$ ,  $p = 1$ ), or Writing 2 ( $V=32$ ,  $p > 0.6$ ).

Figure 1 shows the skill slipping rates across the five datasets. The slipping rates of skills on the Number Line, Writing, and Writing 2 datasets are effectively zero (the max slip rate for any skills in these datasets is  $9 \times 10^{-9}$  percent), which explains why there is no significant difference in model fit for these datasets; i.e., the AFM+S is practically identical to AFM for these datasets. Further, it is likely that there was no difference on student and item cross validation because there was not enough statistical power to detect a difference; i.e., I performed only 1 run of 2-fold cross validation and only two of the five datasets had skills with non-zero slipping rates.

Across all five datasets AFM+S estimates higher skill intercepts ( $V=257.5$ ,  $p < 0.01$ ) and slopes ( $V=117$ ,  $p < 0.01$ ) than AFM, whereas AFM estimates higher student intercepts ( $V=9226$ ,  $p < 0.01$ ) than AFM+S (via a Wilcoxon Signed-Rank test paired by skill and dataset). Note, these results are being primarily driven by the Geometry and Equation Solving datasets because AFM and AFM+S are practically identical on the Number Line, Writing, and Writing 2 datasets.

## 5. DISCUSSION

In general, my results show that AFM+S better fits the data than the AFM model and that there are significant differences in the

parameters estimated by the two models. In particular, the skill intercepts and learning rate estimates from the AFM+S model are higher than those returned by the AFM model. Further, the student intercept estimates from AFM+S are lower than those produced by AFM. These findings suggest that the AFM model might be compensating for skill slipping by adjusting the other parameters. The implication of this finding is that researchers interpreting parameter estimates returned by AFM should be cautious in situations where skill slipping appears to be occurring.

These results also suggest that, at least for these five datasets, the AFM+S model is generally preferable to the AFM model. In situations where no slipping is occurring AFM+S reduces to the AFM model and returns statistically identical model fits. However, when slipping occurs model fit improves with AFM+S.

In conclusion, I have introduced a LearnSphere workflow component and shown how this component can be used to investigate the differences in model fits and parameter estimates of the AFM and AFM+S models. My analysis shows that AFM+S better fits the data than AFM on datasets where slipping occurs and that there are significant differences between the parameter estimates returned by the two models. These results suggest that researchers using the AFM model should consider transitioning to the AFM+S model when they suspect slipping to be occurring. These results also showcase the capabilities of the new LearnSphere AFM+S workflow component.

## 6. ACKNOWLEDGMENTS

We thank Erik Harpstead, Michael Yudelson, and Rony Patel for their thoughts and comments when developing this work. This work was supported in part by the Department of Education (#R305B090023 and #R305B110003) and by the National Science Foundation (#SBE-0836012). Finally, we thank Carnegie Learning and all other data providers for making their data available on DataShop.

## 7. REFERENCES

- [1] Hao Cen, Kenneth R Koedinger, and Brian Junker. 2006. Learning Factors Analysis – A General Method for Cognitive Model Evaluation and Improvement. 164–175.
- [2] Albert T Corbett and John Robert Anderson. 1995. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction* 4, 4: 253–278.
- [3] Christopher J MacLellan, Ran Liu, and Kenneth R Koedinger. 2015. Accounting for Slipping and Other False Negatives in Logistic Models of Student Learning.
- [4] Ran Liu. Personal Communication. 2016.
- [5] Kenneth R Koedinger. Geometry Area 1996-1997. Dataset 76 in DataShop. Retrieved from [pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=76](http://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=76).
- [6] Ritter, S., Anderson, J.R., Koedinger, K.R., & Corbett, A. 2007. The Cognitive Tutor: Applied research in mathematics education. *Psychonomics Bulletin & Review*, 14, 2: 249-255.
- [7] Booth, J., & Ritter, S. 2009. Self Explanation sch\_a3329ee9 Winter 2008 (CL). Dataset 293 in DataShop. Retrieved from [pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=293](http://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=293).
- [8] Derek Lomas. Digital Games for Improving Number Sense – Study 1. Dataset 445 in DataShop. Retrieved from [pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=445](http://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=445).
- [9] Ruth Wylie. IWT Self-Explanation Study 1 (Spring 2009) (tutors only). Dataset 313 in DataShop. Retrieved from [pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=313](http://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=313).
- [10] Ruth Wylie. IWT Self-Explanation Study 2 (Spring 2009) (tutors only). Dataset 372 in DataShop. Retrieved from [pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=372](http://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=372).
- [11] Dietterich, T. G. 1998. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7), 1895–1923. <http://doi.org/10.1162/089976698300017197>.