Dynamic Forecast Combinations

Unlocking FINNs Potential

Advanced Analytics Apprenticeship Program

Michael Buttitta
Shaurya Dubey
Milinda Bhopatkar

Introduction

How FINN combines forecasts today and *how we can do better.*

How FINN Combines Forecast Models Today

User Loads Data

Historical Data uploaded to FINN

Creates Individual and Ensemble Models

Component models used for combination forecasts

Creates New Combined Models

Simple Average of 2-3 component models

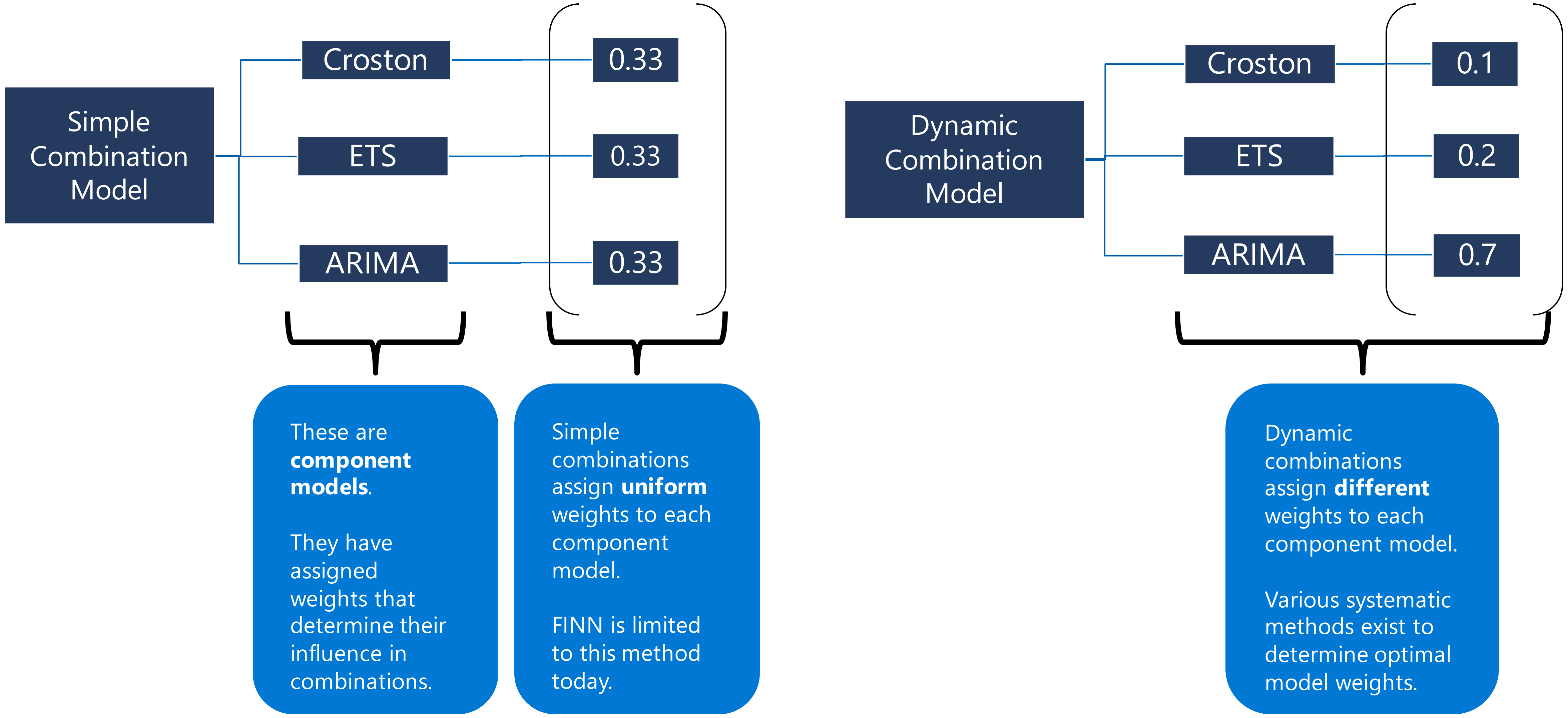
We aim to test dynamic averages in this step

Evaluates MAPE of All Models

Outputs “Best Model”

MAPE: Mean Absolute Percentage Error

Dynamic Combinations



Why Dynamic Forecasts?

Today, FINN is limited to simple model combinations and lacks functionality to place a *customized emphasis* on "useful models"

Dynamic Forecast Models address these shortcomings:

- More robust than counterparts
- Makes FINN more complete
- Lays foundational groundwork for FINN improvements

"All models are wrong, but some are useful."
- George Box, Statistician

How can we implement dynamic forecasting within FINN and measure its performance?

Methodology & Implementation

The ForecastComb Package, implementing with FINN, and experiment design.

The ForecastComb Package

What solutions exist in creating dynamic weights?

ForecastComb offers:

- Widest range of combination methods
- Standardized development framework
- Built-in data processing tools
- Active developer community

ForecastComb Model Combination Methods

Regression	Eigenvector	Simple
• 4 methods • Best when few models outperform the others	• 4 methods • Best when models have similar accuracy	• 7 methods • FINN only applies one simple method

Various Alternatives Exist:

- Prioritized range of methods, ease of use, and support documentation

- BMA
- opera
- forecastHybrid

Applying ForecastComb to FINN Forecasts

How do we combine FINN component forecasts using ForecastComb dynamic weights?

Obtain Individual Models through FINN → Extract Residuals from R Objects (1) → Calculate Component Forecasts (2) → Utilize ForecastComb (3)

1

	.model_id	.model	.model_desc
1	1	*list(pre = list(actions = list(recipe = list(recip [...]*	theta
2	2	*list(pre = list(actions = list(recipe = list(recip [...]*	arima
3	3	*list(pre = list(actions = list(recipe = list(recip [...]*	mars-R1

```r
if(m == 'arima' || m == 'ets'|| m == 'theta'){}
else if(m == 'croston' || m == 'prophet'|| m == 'stlm-arima'|| m == 'stlm-ets'){}
else if(m == 'mars-R1'){
  residuals_vector_model <- modeloutputR1[[2]][[model_id]][["fit"]][["fit"]][["fit"]][["residuals"]]
}
```

2

Residuals = Actuals – Component Forecasts

<u>Component Forecasts</u> = Actuals – Residuals

3

input_data	list [4] (S3: foreccomb)	List of length 4
Actual_Train	double [102] (S3: ts)	9934041 69597284 56427864 15708182 33614784 47754705 ...
Forecasts_Train	double [102 x 3] (S3: mts, ts, mat	79064961 79064961 79064961 79064961 79064961 79064961
nmodels	integer [1]	3
modelnames	character [3]	'mars-R1' 'arima' 'theta'

Run foreccomb object through regression / eigenvector / simple combinations methods

Experiment Design

How do we compare effectiveness of dynamic models to FINNs current functionality?

Do dynamic combinations yield forecasts with lower MAPE's compared to what FINN is capable of today?

1 FINN Maximum Capability

- 23 Individual Models
- Ensemble Models
- Simple Combinations

FINN Backtesting

2 FINN Minimum Component Models

- ARIMA, Mars, Theta
- Simple Combinations

FINN Backtesting

3 ForecastComb Simple Models

- ARIMA, Mars, Theta
- Simple Combinations

Trained only on Full Historical Dataset

4 ForecastComb Dynamic Models

- ARIMA, Mars, Theta
- Dynamic Combinations

Trained only on Full Historical Dataset

Best MAPE	1	2	3	4
Office Revenue				
C+E Revenue				
Surface Revenue				

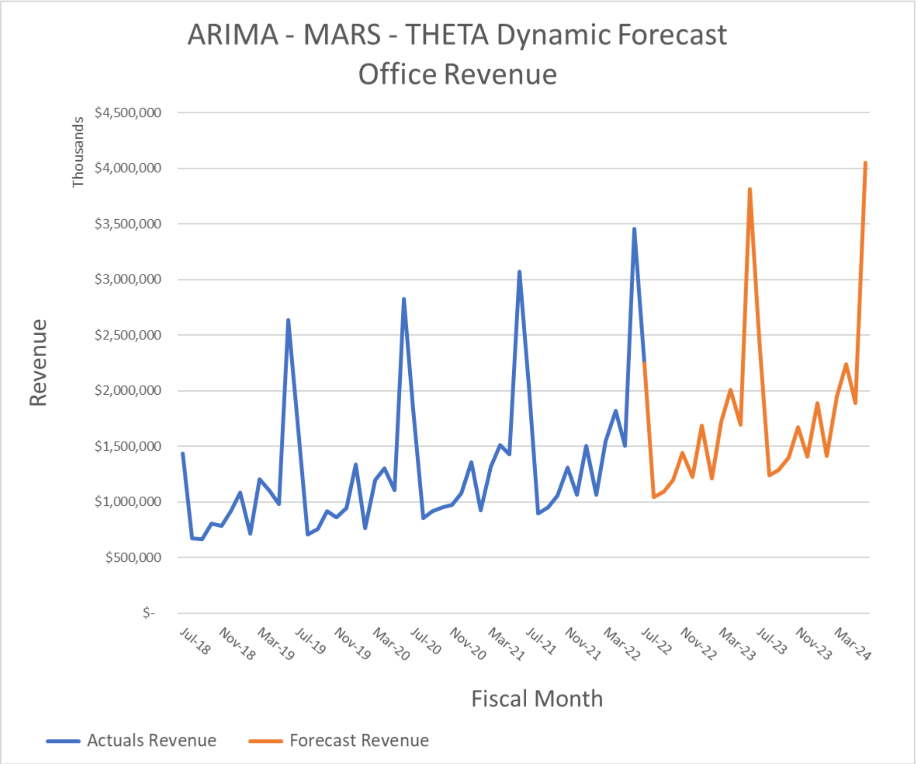
Results & Outlook

Was it worth it?

Dynamic Combination Performance

ForecastComb Dynamic Models beat ForecastComb Simple models; FINN comparison inconclusive.

Best MAPE	FINN Maximum	FINN Comparison	ForeComb Simple	ForeComb Dynamic
Office Revenue	6.5%	6.4%	13.0%	6.9%
C+E Revenue	8.2%	10.7%	13.7%	7.8%
Surface Revenue	15.6%	20.2%	31.0%	19.0%



1. Dynamic models outperformed simple models by +8pts on average
2. Dynamic models outperformed FINN comparison in C+E Revenue by +20%; negligible differences otherwise

Model Weights:

mars-R1	0.14467073
arima	0.94672869
theta	-0.09949517

NOTE:
FINN MAPE's yielded in a back testing process with selection bias (cross-validation).
ForecastComb MAPE's are yielded from a model trained only on all historical data.

Outlook

FINN can gain strength from dynamic forecasts in the quest for a lower MAPE.

Initial tests for dynamic forecasts are promising

- Customized emphasis outperformed uniform emphasis

Learnings:

1. FINN backtesting includes bias for recent observations and immediate forecasts
   - Best Models MAPE not necessarily the lowest MAPE due to this bias
2. Full implementation requires FINN backtesting and access to all component residuals

"...I would think the number of times the best model is a simple model average could be as high as 75-80%"
- Mike Tokic, FD&E

1 FINN Backtesting

Backtest	Train:Test Ratio	Forecast @ T=1	Backtest Weights
1	100/0	$ 455,855	0.4
2	90/10	$ 468,211	0.3
3	70/30	$ 439,130	0.2
4	50/50	$ 488,267	0.05
5	30/70	$ 428,305	0.05

Final Forecast $ 456,460

2 FINN Backtesting + Dynamic Weights

Backtest	Train:Test Ratio	ARIMA	MARS	THETA	Dynamic Fcst. @ T=1	Backtest Weights
1	100/0	0.3	0.4	0.3	$ 458,952	0.4
2	90/10	0.2	0.3	0.5	$ 475,222	0.3
3	70/30	0.8	0.1	0.1	$ 473,987	0.2
4	50/50	0.7	0.2	0.1	$ 480,626	0.05
5	30/70	0.6	0.3	0.1	$ 426,591	0.05

Final Forecast $ 466,306