

Measurement Report

Jason Ma

November 2018

1 Introduction

This report contains a summary of the rate of inconsistency between various data sources and End of Day EMA. Ideally, if a signal is reported in a data source, then that signal should be reflected in the End of Day EMA file. We calculate the percentage of signals that were not reported. We report two measurements: percentage by entry and percentage by date (there could be multiple signals in one day, and we count them all together as one error).

2 Contingent vs. EOD EMA

	Original	Alternative	Back-up
% by entry	0.156	0.131	0.114
% by date	0.1	0.115	0.0757

Table 1: Contingent vs. EOD EMA

3 Random vs. EOD EMA

	Original	Alternative	Back-up
% by entry	0.207	0.237	0.171
% by date	0.159	0.158	0.133

Table 2: Random vs. EOD EMA

4 Puff-Episode vs. EOD EMA

	Original	Alternative	Back-up
% by entry	0.298	0.327	0.294
% by date	0.121	0.127	0.129

Table 3: Puff-Episode vs. EOD EMA

5 Summary

The error rates for contingent and random EMA float around 15%, though contingent EMA has slightly lower inconsistency. It appears that the back-up data has lower inconsistency rate across various data sources. Finally, the error rate for Puff-Episode is inflated because not all puff episodes are actually smoking episodes that were intended to be reported in EOD EMA.