Sam Pfeiffer

MATH 9898 – Big Data in Finance

Assignment C

March 22, 2015

For this assignment, I created two programs that calculate the mean, volatility, and drift of a time series.

To generate the data, I used the daily fed fund rate. I generated intraday data by using Brownian Bridge with the volatility of the synthetic data matching that of the input data.

The parameters for the programs are in *parameters.txt*. Note that to generate a ~45GB data file, the “average time between ticks” parameter should be set to 93.

To generate the data run:

$ make generate\_data.out

$ ./generate\_data.out

I then wrote the program using both MPI and hadoop.

To compile the MPI program, run

$ make mpi\_mapreduce.out

To run it,

$ mpirun –np 10 ./mpi\_mapreduce.out

Note that to build the *generate\_data.cpp* and *mpi\_mapreduce.cpp* programs, you need a g++ compiler capable of using c++11.

To run the Hadoop program:

$ make hadoop\_mapreduce

To run it,

$ mapred pipes -conf hadoop\_mapreduce.xml -input in-dir -output out-dir

To read the output,

$ hadoop fs -cat out-dir/part-00000