**Codes and Data Readme**

**twostate.m**

This is the code in which we generated two state example used in the presentation. This code helped us in making sure that we are doing the right thing. It also helped us in understanding many aspects of the filtering as well as interpretations of our results.

**df.mat**

Df is the re-organized default data. Its index is the number of days since the beginning of the sample period, and its value is the number of defaults in that specific day. We used excel to get the days and thus did not include a matlab file to get these data.

**main.m**

This is the main code that calls the functions to train different models for the actual default data. It will output a figure with all the essential results, such as the default rate matrix, filter state at each time point, as well as the learning curve. This file also uses the log-likelihood values from the tested models to calculate the BIC values. Results are plotted and saved.

**ergodic\_leftright.mat**

This is the results from main.m.

**trainHMM.m**

This is the code that actually does the training of model given the data and initial specification of the data.

**hmmtrain.m**

This is the matlab built-in HMM training file. It used a scaling method in order to deal with numerical underflow. We also modified it in order to train our non-communicating models. This is done by setting the corresponding transition matrix element to be zero.

**get\_parameters\_outofsampletest.m**

This is the code we used to train our model using different portion of the data. It outputs necessary variables for us in the later prediction procedures. In fact this code generated two-state for prediction parameters, but due to its poor performance in the in-sample test, we decide not to use it further.

**outofsample\_pred\_2.m**

This is the code that used the parameters to generate the default events in the next year. It calls hmmgenerate\_edit.m. It also calculate the normalized probability density function of the predict default events in the future year. Results are plotted and saved.

**hmmgenerate\_edit.m**

This is the code in which we can input the transition matrix, default rate matrix and the initial state to generate one realization with length L. One important thing that we did not do is that we should write a code that can take a history of the previous data as input and then generate one realization. This is a major reason why our out-of-sample test did not do well.

**playpred.m**

This is the code we used to get the quantiles from the normalized pdf of predicted-default events.