Regime Switching and Technical Trading with Dynamic Bayesian Networks in High-Frequency Stock Markets

Luis Damiano, Brian Peterson, Michael Weylandt

2017-08-26

Contents

This work aims at … .

The authors acknowledge Google for financial support via the Google Summer of Code 2017 program.

1 Motivation

2 Hierarchical Hidden Markov Models

The Hierarchical Hidden Markov Model (HHMM) is a recursive hierarchical generalization of the HMM that provides a systematic unsupervised approach for complex multi-scale structure. The model is motivated by the multiplicity of length scales and the different stochastic levels (recursive nature) present in some sequences. Additionally, it infers correlated observations over long periods via higher levels of hierarchy.

The model structure is fairly general and allows an arbitrary number of activations of its submodels. The multi-resolution structure is handled by temporal experts[1] of different time scales.

[1] In Machine Learning terminology, a problem is divided into homogeneous regions addressed by an expert submodel. A gating network or function decides which expert to use for each input region.

2.1 Model specification

HHMM are structured multi-level stochastic processes that generalize HHM by making each of the hidden states an autonomous probabilistic model. There are two kinds of states: internal states are HHMM that emit sequences by a recursive activation of one of the substates, while production states generate an output symbol according to the probability distribution of the set of output symbols.

Hidden dynamics are lead by transitions. Vertical transitions involve the activation of a substate by an internal state, they may include further vertical transitions to lower level states. Once completed, they return the control to the state that originated the recursive activation chain. Then, a horizontal transition is performed. Its state transition within the same level.

A HHMM can be represented as a standard single level HMM whose states are the production states of the corresponding HHMM with a fully connected structure, i.e. there is a non-zero probability of tranisition from any state to any other state. This equivalent new model lacks the multi-level structure.

Let $z_t^d = i$ be the state of an HHMM at the step t, where $i \in \{1, \ldots, |z^d|\}$ is the state index, $|z^d|$ is the number of possible steps within the d-th level and $d \in \{1, \ldots, D\}$ is the hierarchy index taking values $d = 1$ for the root state, $d = \{2, \ldots, ..., D-1\}$ for the remaining internal states and $d = D$ for the production states.

In addition to its structure, the model is characterized by the state transition probability between the internal states and the output distribution of the production states. For each internal state z_t^d for $d \in \{1, \ldots, D-1\}$, there is a state transition probability matrix $\boldsymbol{A}^d$ with elements $A_{ij}^d = p(z_t^{d+1} = j | z_t^{d+1} = j)$ is the probability of a horizontal transition from the i-th state to the j-th state within the level d. Similarly, there is the initial distribution vector over the substates $\boldsymbol{\pi}^d$ with elements $\pi_j^d = p(z_t^{d+1} = j | z_t^d)$ for $d \in \{1, \ldots, D-1\}$. Finally, each production state z_t^D is parametrized by the output parameter vector $\boldsymbol{\theta}_o^i$ whose form depends on the specification of the observation model $p(\boldsymbol{x}_t | z_t^D = j, \boldsymbol{\theta}_o^j)$ corresponding to the j-th production state.

2.2 Generative model

The root node initiates a stochastic sequence generation. An observation for the first step in the sequence t is generated by drawing at random one of the possible substates according to the initial state distribution $\boldsymbol{\pi}^1$. To replicate the recursive activation process, for each internal state entered z_t^d one of the substates is randomly chosen according to the corresponding initial probability vector $\boldsymbol{\pi}^d$. When an internal state transitions to a production state $z_t^D = j$, a single observation is generated according to the state output parameter vector $\boldsymbol{\theta}_o^j$. Control returns to the internal state that lead to the current production state z_t^{D-1}, which in turns selects the next state in the same level according to transition matrix $\boldsymbol{A}^{D-1}$.

Save for the top, each level $d \in \{2, \ldots, D\}$ has a final state that terminates the stochastic state activation process and returns the control to the parent state of the whole hierarchy. The generation of the observation sequence is completed when control of all the recursive activations returns to the root state.

2.3 Parameter estimation

The parameters of the models are $\boldsymbol{\theta} = \left\{ \left\{ \boldsymbol{A}^d \right\}_{d \in \{1, \ldots, D-1\}}, \left\{ \boldsymbol{\pi}^d \right\}_{d \in \{1, \ldots, D-1\}}, \{\boldsymbol{\theta}_o\} \right\}$. The form of $\boldsymbol{\theta}_o$ depends on the specification of the production states. We refer the read to Fine and Singer (1998) for a detailed treatment of estimation and inference procedures.

3 Regime Switching and Technical Trading with Dynamic Bayesian Networks in High-Frequency Stock Markets

3.1 Preamble

...

3.2 Feature extraction

3.2.1 Input series

Tick series are a sequence of triples $\{y_k\}$ with $y_k = (t_k, p_k, v_k)$, where $t_k \leq t_{k+1}$ is the time stamp in seconds, p_k is the trade price and v_k is the trade volume. The sequence is ordered by the occurence of trades. There can be more than one trade within a second.

Following Tayal (2009), who in turns drew inspiration from the technical analysis techniques proposed by Ord (2008), we derive a zig-zag sequence that captures the bid-ask bounce $\{z_k\}$ with $z_k = (i_n, j_n, e_n, \phi_n)$, where $i_n \leq i_j$ are indices to the tick series representing the starting and ending point of the extrema, $e_n = p_k \; \forall \; k : i_n \leq k \leq j_n$ is the price at the local extrema, and ϕ_m measures the average volume per second during the zig-zag leg ending at e_n:

$$\phi_n = \frac{1}{t_{j_n} - t_{i_{n-1}} + 1} \sum_{k=i_{n-1}}^{j_n} v_k.$$

We note that $p_{i_n} < e_n < p_{j_n+1}$ for local maxima and $p_{i_n} > e_n > p_{j_n+1}$ for local minima. The average volume, which includes the end-point extrema, is normalized by $t_{j_n} - t_{i_{n-1}} + 1$ to avoid division by zero when the zig-zag leg occurs within the same time period. Most importantly, we underline that the n-th zig-zag point z_n is realized only after observing the $(j_n + 1)$-th tick point y_{j_n+1}. Failing to consider the one tick lag between leg completion and the time of detection would cause look-ahead bias in the out of sample forecasts.

3.2.2 Processing rules

Discrete features are created based on the zig-zag series $\{z_n\}$. We first create an auxiliary series $\{O_n\}$ with $O_n = (f_n^0, f_n^1, f_n^2)$, where f_n^0 represents the direction of the zig-zag, f_n^1 indicates the existence of a trend and f_n^2 indicates whether average volume increased or decreased.

Formally,

$$f_n^0 = \begin{cases} +1 & \text{if } e_n \text{ is a local maximum (positive zig-zag leg)} \\ -1 & \text{if } e_n \text{ is a local minimum (negative zig-zag leg),} \end{cases}$$

and

$$f_n^1 = \begin{cases} +1 & \text{if } e_{n-4} < e_{n-2} < e_n \wedge e_{n-3} < e_{n-1} \text{ (up-trend)} \\ -1 & \text{if } e_{n-4} > e_{n-2} > e_n \wedge e_{n-3} > e_{n-1} \text{ (down-trend)} \\ 0 & \text{otherwise (no trend).} \end{cases}$$

For the third indicator function, we compute the average volume ratios,

$$\nu_n^1 = \frac{\phi_n}{\phi_{n-1}}, \quad \nu_n^2 = \frac{\phi_n}{\phi_{n-2}}, \quad \nu_n^1 = \frac{\phi_{n-1}}{\phi_{n-2}},$$

we transform the ratios into a discrete variable using an arbitrary threshold α,

$$\tilde{\nu}_n^j = \begin{cases} +1 & \text{if } \nu_n^j - 1 > \alpha \\ -1 & \text{if } 1 - \nu_n^j > \alpha \\ 0 & \text{if } |\nu_n^j - 1| \leq \alpha, \end{cases}$$

and we finallly define

$$f_n^2 = \begin{cases} +1 & \text{if } \tilde{\nu}_n^1 = 1, \tilde{\nu}_n^2 > -1, \tilde{\nu}_n^3 < 1 \text{ (volume strengthens)} \\ -1 & \text{if } \tilde{\nu}_n^1 = -1, \tilde{\nu}_n^2 < -1, \tilde{\nu}_n^3 > -1 \text{ (volume weakens)} \\ 0 & \text{otherwise (volume is indeterminant).} \end{cases}$$

The features or legs $\boldsymbol{D} = \{D_1, \dots, D_9\}$, $\boldsymbol{U} = \{U_1, \dots, U_9\}$ are then created using the Table 1.

	Zig-zag direction	Price trend	Change in volume	Market State
D_1	Down -1	Up +1	Weak -1	Bull
D_2	Down -1	Down -1	Weak -1	Bull
D_3	Down -1	Up +1	Intederminant 0	Bull
D_4	Down -1	No trend 0	Weak -1	Bull
D_5	Down -1	No trend 0	Intederminant 0	Local volatility
D_6	Down -1	No trend 0	Strong +1	Bear
D_7	Down -1	Down -1	Intederminant 0	Bear
D_8	Down -1	Up +1	Strong +1	Bear
D_9	Down -1	Down -1	Strong +1	Bear
U_1	Up +1	Up +1	Strong +1	Bull
U_2	Up +1	Down -1	Strong +1	Bull
U_3	Up +1	Up +1	Intederminant 0	Bull
U_4	Up +1	No trend 0	Strong +1	Bull
U_5	Up +1	No trend 0	Intederminant 0	Local volatility
U_6	Up +1	No trend 0	Weak -1	Bear
U_7	Up +1	Down -1	Intederminant 0	Bear
U_8	Up +1	Up +1	Weak -1	Bear
U_9	Up +1	Down -1	Weak -1	Bear

Table 1: Feature space.

Defining appopriate rules that capture trade volume dynamics and identify trends in volume despite high-frequency noise is the most challenging aspect of this design. Wisebourt (2011) proposes a modification for the feature extraction procedure. By computing the spread between the Volume Weighted Average Prices of the bid and the ask, he designs a book imbalance metric that describes the state of the order book at any given moment in time. In turns, Sandoval and Hernández (2015) applies wavelets over two simple-smoothed exponential distance-weighted average volume series to measure trade volume concentration in both sides of the book.

3.3 Model

We adhere to the methodology proposed in the original work as much as possible. We set up a HHMM to learn the sequence of discrete features extracted from a high-frequency time series of stock prices and traded volume. The figure below summarises the model structure in the form of a Dynamic Bayesian Network.

The graph starts with a root node z^0 that has two top-level children z_1^1 and z_2^1 representing bullish markets (or runs) and bearish markets (or reversals). The specifications do not pose any constraints to determine beforehand which node takes each of the possible two meanings. In consequence, latent states need to be labeled after the learning stage based on sample characteristics such as mean returns. Althought the original author does not mention this possibility, prior information, like parameter ordering, could be embedded to break symmetry and mitigate eventual identification issues.

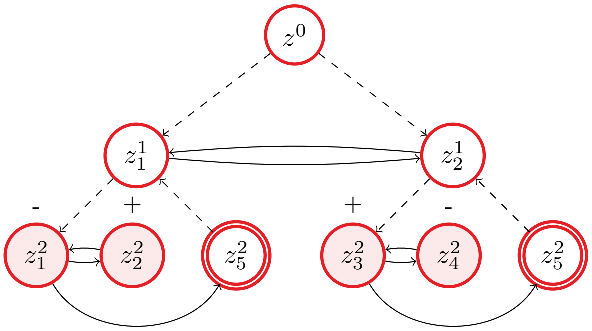

Figure 1: Hierarchical Hidden Markov Model for price and volumen proposed by @tayal2009regime.

Each top node activates a probabilistic HMM with two latent states for negative and positive zig-zag legs. The sub-model activated by the node z_1^1 always starts with the node z_1^2 producing an observation from the distribution of negative zig-zag $\boldsymbol{D} = \{D_1, \dots, D_9\}$. Next, it transitions to (a) the node z_2^2 producing a positive zig-zag leg $\boldsymbol{U} = \{U_1, \dots, U_9\}$, (b) the end node which may, in turn, transition back to z_1^1 emitting a new negative zig-zag leg, or (c) the end node and switches to the second sub-model, landing on node z_3^2 and producing a positive zig-zag leg. Restricting transitions to this limited set of movements force alternation between positive and negative zig-zag legs, thus guaranteeing that all possible observation sequences are well behaved. The sub-model belonging to z_2^1 has similar but symmetrical behaviour. These two inner models are conditionally independent, an advantage we will take for computational time.

The persistence of price trends vary according to the resolution of the dataset. In low frequency contexts, financial analysts will frequently define short and long term trends based on timespans involving weeks and months of observations. In high frequency trading, however, trends may last from seconds to hours. The main idea behind the conception of the hierarchical model is that top levels may group a sequence of price movements forming a general trend, whereas the bottom nodes allow zig-zags to represent micro trends that may deviate shortly from the main drift. That is, the model accommodates for possibly unequally probable negative price movements in an upwards market as well as positive short trends in a price reversal.

Theoretically, all HHMM can be expressed as an equivalent HMM with possibly sparse initial probability vector and transition matrix. Although learning from this representation may prove less efficient in terms of computational complexity, the relatively simple structure and many restrictions of the model under study make estimation and inference feasible. The equivalent "expanded" HMM has $K = 4$ states, the following K-sized initial probability vector

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_1 & 0 & 1 - \pi_1 & 0 \end{bmatrix}$$

and the $K \times K$ transition matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & a_{12} & 1 - a_{11} & 0 \\ 1 & 0 & 0 & 0 \\ a_{31} & 0 & 0 & 1 - a_{31} \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

where each element a_{ij} represents the probability of transitioning from the hidden state i (row) to j (column) in one step. The matrix is sparse with zeros representing nodes with no direct connections. Since the initial probability vector and the rows of the transition matrix must sum to 1, hidden dynamics are governed by only three free parameters.

Production nodes emit one observation from the finite sets of possible outputs $\boldsymbol{O}_D$ (for z_1^2 and z_4^2) and $\boldsymbol{O}_U$ (for z_2^2 and z_3^2). Conditional probability distributions are given by output probability vectors of length $L = 9$ subject to sum-to-one constraints:

$$\begin{aligned}
\boldsymbol{B}^1 &= p(\boldsymbol{x}_t|z_1^2) = \left[b_{D_1}^1, \dots, b_{D_9}^1\right], \\
\boldsymbol{B}^2 &= p(\boldsymbol{x}_t|z_2^2) = \left[b_{U_1}^2, \dots, b_{U_9}^2\right], \\
\boldsymbol{B}^3 &= p(\boldsymbol{x}_t|z_3^2) = \left[b_{U_1}^3, \dots, b_{U_9}^3\right], \\
\boldsymbol{B}^4 &= p(\boldsymbol{x}_t|z_4^2) = \left[b_{D_1}^4, \dots, b_{D_9}^4\right].
\end{aligned}$$

The observation model has 32 free parameters.

On the whole, the parameter vector for the HMM representation reduces to vector of size 35,

$$\boldsymbol{\theta} = (\pi_1, a_{12}, a_{31}, b_{D_l}^1, b_{U_l}^2, b_{U_l}^3, b_{D_l}^4),$$

with $l \in \{1, \dots, L-1\}$.

3.4 Dataset

The original work presents results for both simulated and real data. The latter is based on historical high-frequency time series for the 60 stocks listed in the S&P/TSE60 index. The dataset consists of all 22 business days of May 2007. The author excludes three days due to significant errors without disclosing the exact dates. We confirm that our results are consistent with the original work by focusing on GoldCorp Inc (TSE:G), the only series described exhaustively in the original work.

Note that we do not model prices directly. Instead, non-linear transforms are applied on the price and traded volume high-frequency series to produce the sequence of observations fed to the proposed model.

3.5 Methodology

Model parameters are estimated on a rolling window with five days each. Since top nodes are symmetrical, states are labeled ex-post based on the order of the in-sample mean of the percentage change in the initial and final price before the top level state switch. The state with larger and smaller returns are marked as bullish (a run) and bearish (a reversal) respectively.

After learning and labelling, two out of sample inference procedures are run on the sixth day. First, offline smoothing infers the hidden state at time t based on the full evidence of the sixth day. Although this quantity is not useful for trading because of its look-ahead bias, it provides an upper bound benchmark for the model. Second, online filtering is used as a trading rule.

Most of the diagnostics are based on trade returns. For the l-th top-level state switch, the percentage return is defined as

$$R_l = \frac{p_l^e - p_l^s}{p_l^s},$$

where p_l^s and p_l^e are the price at the start and end of the switch.

The information content of learning the top-level state is assessed by comparing the unconditional empirical distribution of trade returns versus their empirical distribution conditioned on the top-level state. Additionaly, regime return characteristics are validated: mean trade returns are expected to be higher for bullish regimes compared to bearish regimes, and they are expected to be positive and negative for runs and reversals respectively. In the original work, most of these analysis are run both in-sample and out-of-sample. We focus on out-of-sample results only.

Finally, a trading strategy is tested. After a zig-zag leg is completed, the trading system buys one unit every time the top-level state switches to bullish (a run) and sells one unit every time it switches to bearish (a reveral). As an addition to Tayal (2009), where trades are executed one tick after the zig-zag leg is observed to ensure there is no lock-ahead bias, we investigate the decay of the strategy performance for longer lags.

3.6 GoldCorp Inc (TSE:G)

We present an in-depth study of one stock to assess the strengths and weaknesses of the model. Using the tick-by-tick series of GoldCorp Inc (TSE:G), we split our dataset in training (2007-05-04 to 2007-05-10 - five trading days and ticks) and test (2007-05-11 - ticks) sets. Next, we run our procedure in a walking forward fashion for this stock as well as others and present some summary statistics for the performance of the strategy.

3.6.1 Data exploration

We center our attention on the sequence of trades, disregarding possibly valuable information from the bid and ask series. Future research may employ such information to improve model predictability. We start by extracting the features using the procedure detailed above. We set the threshold for the change in volume indicator variable in $\alpha = 0.25$ as suggested by the author of the original work. In-sample dataset reduced from trade records to 8386 zig-zags.

Apart from the zig-zags themselves, local extrema are interesting on their own. Although we could not gain insight by visually inspecting other intermediate indicators such as the trend f_1, further research may find value in them.

```
## [1] 1
```

```
## [1] 1
```

```
## Warning in par(cex = 0.7, cex.names = 0.7, cex.axis = 0.7, cex.lab = 0.7, :
## "cex.names" is not a graphical parameter
```

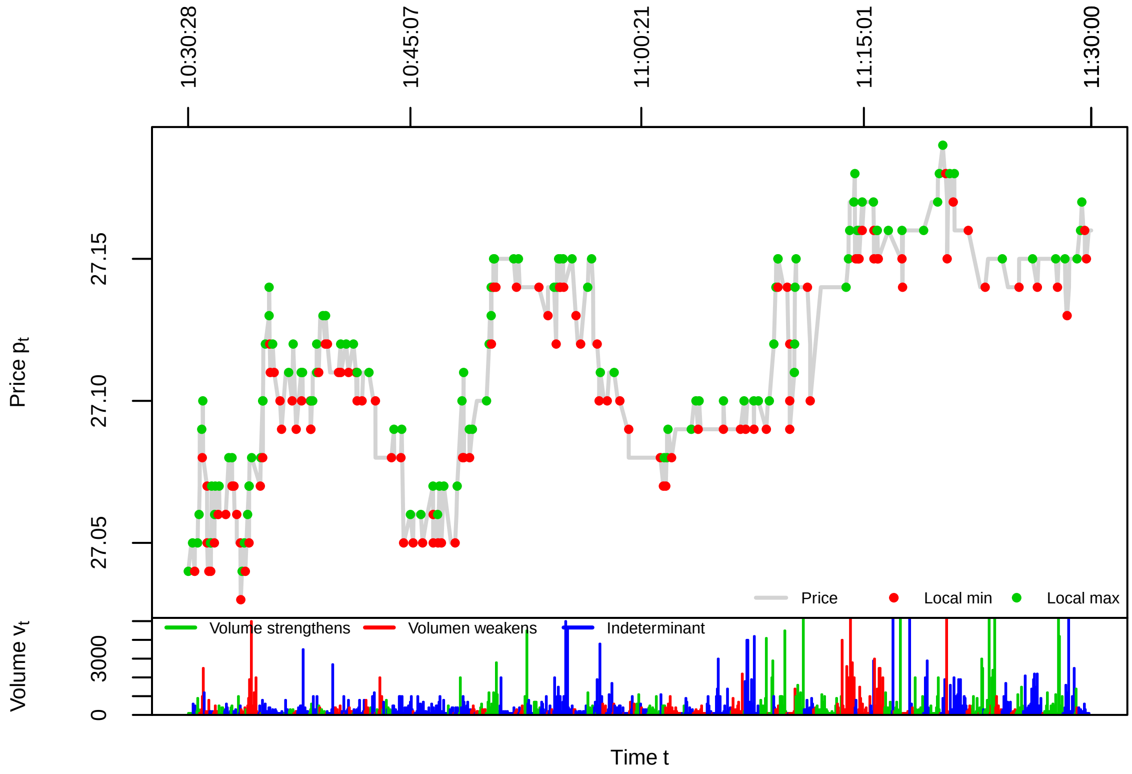

Figure 2: Local extrema detected in TSE:G 2007-05-11 10:30:00/2007-05-11 11:30:00.

```
## [1] 0.7
```

```
## [1] 0.7
```

```
## [1] TRUE
```

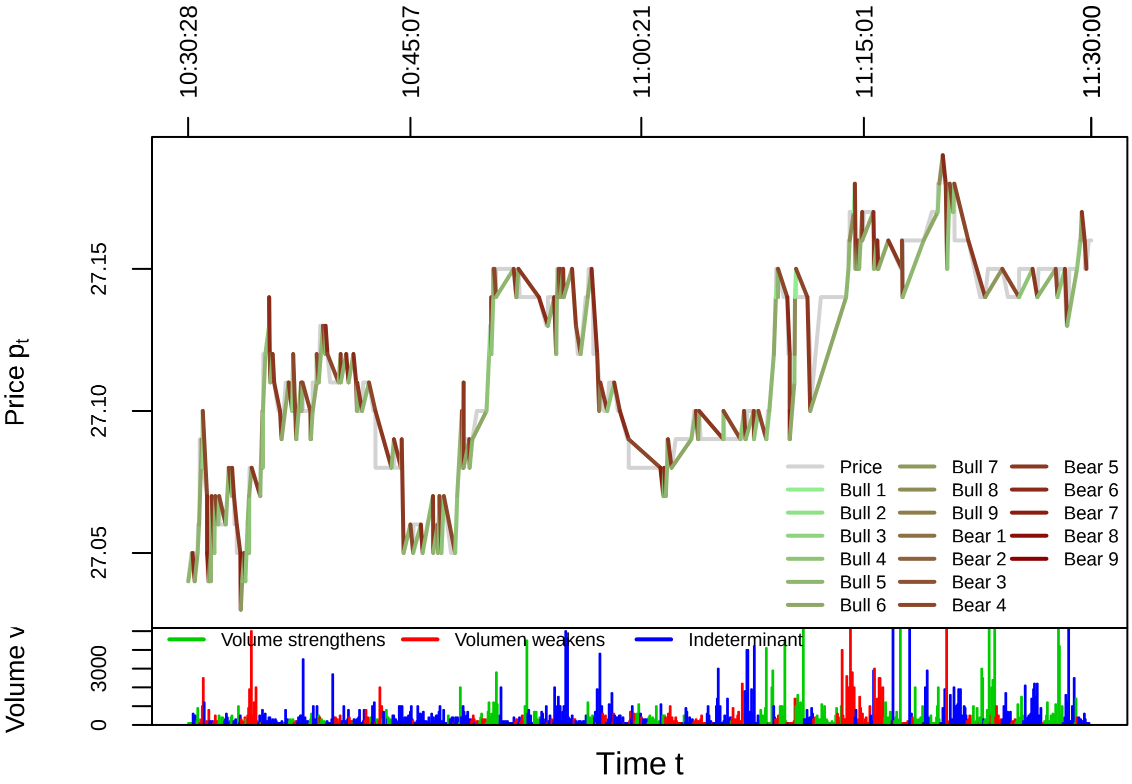

Figure 3: Features extracted from TSE:G 2007-05-11 10:30:00/2007-05-11 11:30:00.

3.6.2 Estimation

We present a summary of the estimates for the model parameters in 2. The top nodes show some persistence since the estimated transition probabilities inside $1 -$
pi_1 and
pi_{31} are close to 1. Extreme features are less likely in all states, which is consistent with our empirical observations. Features D_5 and U_5 are frequent in both regimens. Features $\{D_1, \ldots D_4\}$ are more frequent on the the most frequent positive and negative legs, are linked to local volatility. ** The weakness**.

```
## $xtable.tabular.environment
## [1] "tabularx"
```

Table 2:

Parameter	Mean	Std. Deviation	$q_{10\%}$	$q_{50\%}$	$q_{90\%}$
π_1	0.5067	0.282	0.0956	0.5162	0.8931
$1 - \pi_1$	0.4933	0.282	0.1069	0.4838	0.9044
a_{12}	0.4559	0.1089	0.3085	0.47	0.579
$1 - a_{12}$	0.5441	0.1089	0.421	0.53	0.6915
a_{21}	1	0	1	1	1
a_{31}	0.0914	0.0623	0.0171	0.0836	0.1717
$1 - a_{31}$	0.9086	0.0623	0.8283	0.9164	0.9829
a_{43}	1	0	1	1	1

Parameter	Mean	Std. Deviation	$q_{10\%}$	$q_{50\%}$	$q_{90\%}$
U_1	0.0073	0.0057	7e-04	0.0064	0.0152
U_2	0.0196	0.0047	0.0145	0.0188	0.0263
U_3	0.0072	0.0081	6e-04	0.0047	0.017
U_4	0.3419	0.0208	0.314	0.3436	0.3665
U_5	0.2231	0.0262	0.1873	0.2253	0.2561

Parameter	Mean	Std. Deviation	$q_{10\%}$	$q_{50\%}$	$q_{90\%}$
U_6	0.3531	0.0174	0.3305	0.3549	0.3719
U_7	0.0264	0.0053	0.0205	0.0263	0.0333
U_8	0.0052	0.0043	6e-04	0.0042	0.0109
U_9	0.0161	0.0046	0.0109	0.0155	0.0226
D_1	0.0017	0.0018	2e-04	0.0011	0.004
D_2	0.0226	0.0053	0.0162	0.0223	0.0297
D_3	0.0017	0.0017	1e-04	0.0012	0.0039
D_4	0.0484	0.0333	0.0062	0.046	0.0918
D_5	0.8008	0.0403	0.7497	0.8026	0.8528
D_6	0.0236	0.0174	0.005	0.0203	0.0469
D_7	0.0796	0.008	0.0698	0.0798	0.0891
D_8	0.0011	0.001	1e-04	9e-04	0.0023
D_9	0.0205	0.0045	0.015	0.0204	0.0262
U_1	0.0105	0.0038	0.0062	0.0106	0.0153
U_2	0.0023	0.0022	3e-04	0.0018	0.0051
U_3	0.0251	0.0052	0.0187	0.0251	0.0314
U_4	0.3554	0.0313	0.3159	0.3549	0.3936
U_5	0.1986	0.0363	0.1525	0.1982	0.2421
U_6	0.3884	0.0198	0.3645	0.3886	0.4154
U_7	0.0023	0.0022	4e-04	0.0017	0.0054
U_8	0.0155	0.0038	0.0111	0.015	0.0205
U_9	0.0018	0.0019	2e-04	0.0013	0.0043
D_1	0.0165	0.0066	0.007	0.017	0.0251
D_2	9e-04	9e-04	1e-04	6e-04	0.002
D_3	0.0558	0.011	0.0419	0.0574	0.0677
D_4	0.0212	0.0158	0.0044	0.0185	0.0452
D_5	0.8793	0.0237	0.8476	0.8803	0.9057
D_6	0.0132	0.0122	0.0017	0.0088	0.0314
D_7	0.0019	0.0017	3e-04	0.0014	0.0039
D_8	0.0098	0.0052	0.0028	0.0098	0.0168
D_9	0.0014	0.0014	1e-04	9e-04	0.0032

3.6.3 Convergence

Figure 4 illustrates the trace plot of some arbitrary parameters as well as some diagnostic measures. In general terms, mixing and convergence to the stationary distribution is acceptable. The shrink factor of (Gelman and Rubin 1992), close to 1 for all hidden quantities, indicates an adequate degree of convergence. Sampling is efficient as signaled by effective sample size ratios near to 1 and Monte Carlo Standar Error to posterior standard deviation ratios well below 10%.

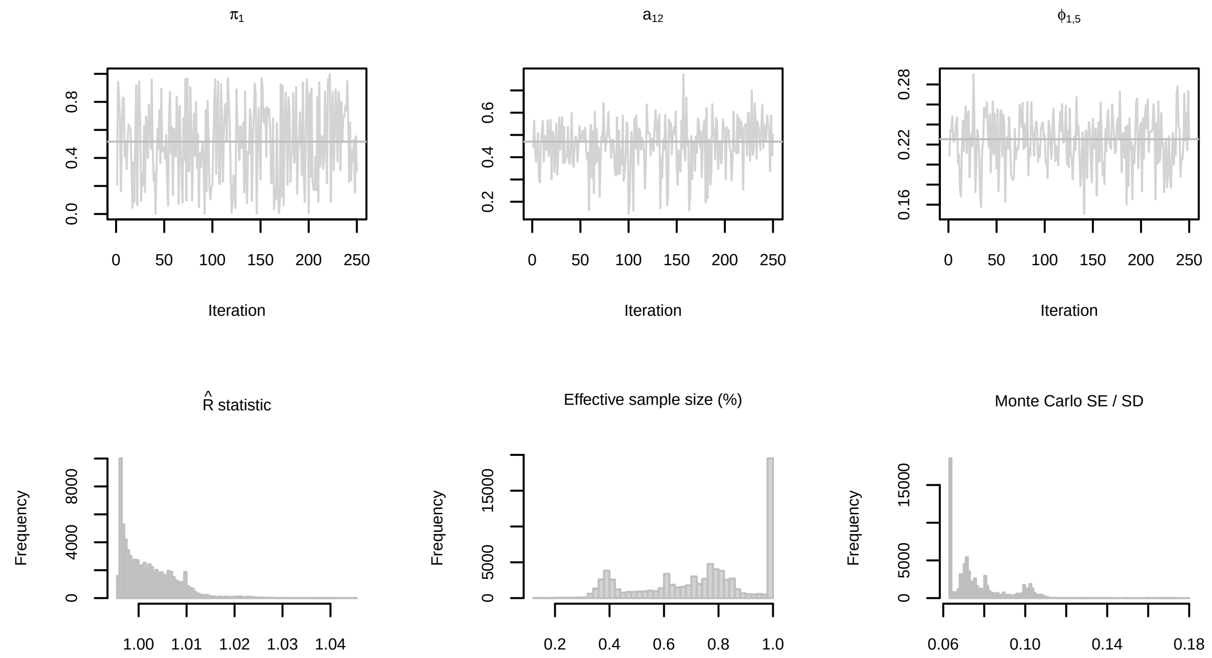

Figure 4: Traceplot of some arbitrarily selected parameters and histograms of diagnostic measures. Mixing, convergence to the stationary distribution and sampling efficiency are acceptable.

Although re-estating the original Hierarchical Hidden Markov Model into a Hidden Markov Model may increase time and memory complexity, the new model becomes significantly easier to program and convergence. A full list of our results, including convergence statistics such as $\hat{R}$ and the effective sample size, is included in the appendix.

3.6.4 State probability

The forward algorithm allows us to calculate the filtered belief state: the probability that an observation at time t was emitted by one of the possible four states (bottom-nodes) given the evidence avaliable up to t. We assign each observation to the emission state with largest filtered probability and, by the definition of the hierarchical model, they become naturally linked to one of the two possible top states (bears or bulls).

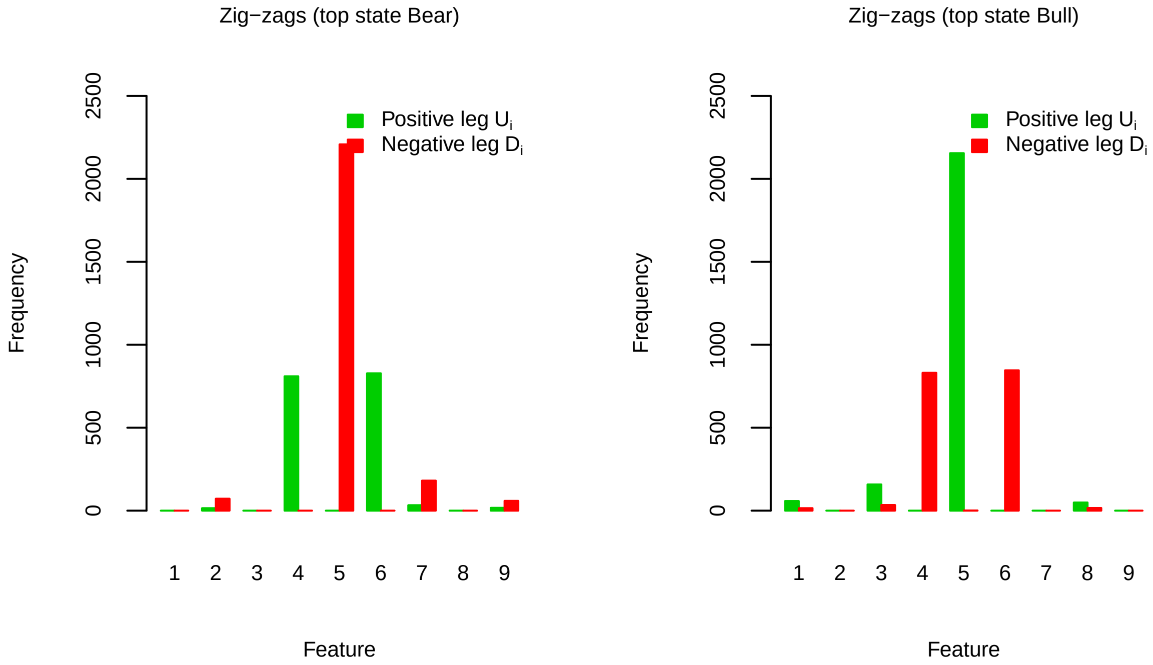

Figure 5: Distribution of observed features conditional on the estimated hidden regime (top node).

The following table provides some summary statistics for the returns of the observations classified in each state. The structure of the top nodes are symmetrical and they do not have an a priori order. We label them as runs and reversals according to the expected trade returns observed in-sample.

Top state	Mean	Std. dev.	Skewness	Kurtosis	$q_{25}\%$	Median	$q_{75}\%$	Mean length	Median length
Bear	-0.0094	0.0648	7.9684	237.4565	-0.036	0	0	10.6215	6
Bull	0.0049	0.0515	0.9683	9.367	0	0	0.0357	10.1696	6
Unconditional	-0.0022	0.0589	5.5346	173.8077	-0.0357	0	0	10.3955	6

Bear and bull top states have negative and positive mean respectively by construction. As it is also visible in Figure 6, a positive skewness coefficient for all states indicates that negative returns tend to overweight positive returns for this specific stock during the five day long in-sample dataset. However, bear markets have a marked skew towards negative returns, becomes more risky in term of extreme events (kurtosis) and trades in bear markets tend to last one tick more than bulls.

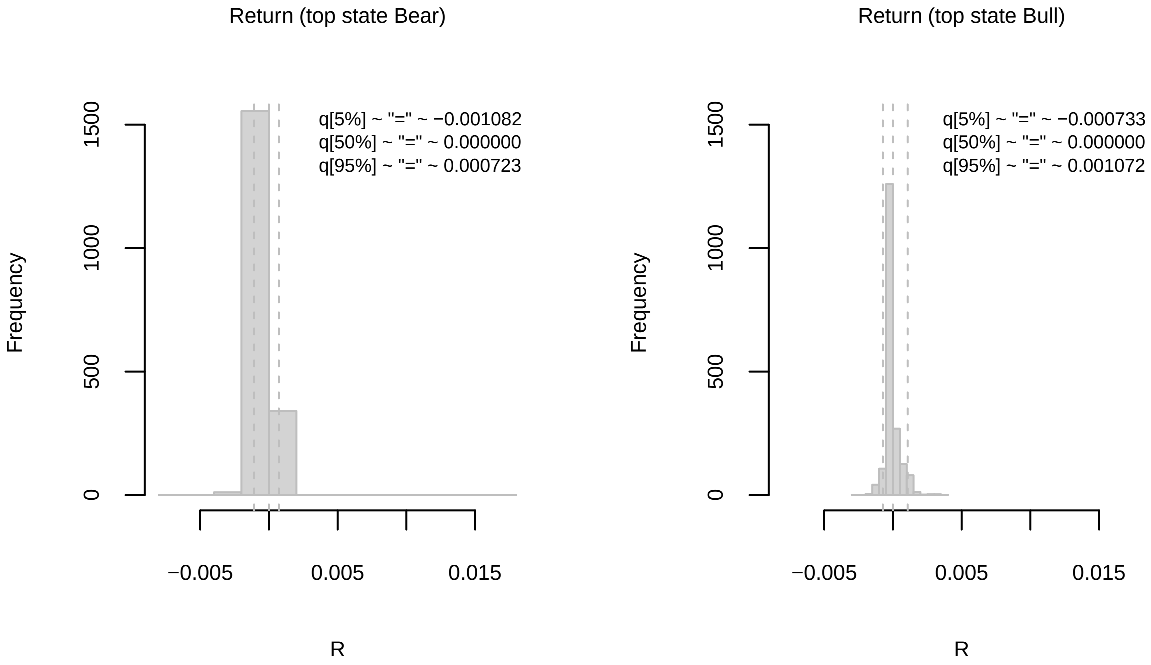

Figure 6: Conditional distribution of trade returns given the estimated top state in TSE:G 2007-05-04 09:30:00/2007-05-10 16:30:00.

We remark that trade returns originated in different stocks or timespans do not necessarily share these characteristics. Location, dispersion, symmetry and shape of the returns vary along stocks and day of analysis.

3.6.5 Fitted output

As mentioned in Section 3.3, the model assumes that outputs are emitted by one of four possible bottom nodes. By definition, negative legs belong to either z_1^2 or z_4^2 while positive legs belong to z_2^2 or z_3^2. Once the parameters were estimated, states z_1^2 and z_2^2 are labeled as bears while z_3^2 and z_4^2 are marked as bulls according to the mean trade return of the observations included in each of them.

Bullish states allow for negative zig-zags and bearish states allow for positive zig-zags as long as the trade volume is indeterminant or weak. The results of the classification are summarized in Table **??**. As expected, the bullish top-node capture positive movements due to local volatility as well as downward movements with weak volume.

Our sampler, tho, maps each of the 18 possible features into one state. For example, observations from feature 3 (ie U3) are all mapped into state 4. In a sense, this adds very little information since imputation is almost deterministic once the parameters are estimated. This is possibly a weakness in the model.

Top state	Emission state	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
Bull	z_1^2	0	15	0	810	0	828	33	0	17	0	0	0	0	0	0	0	0	0
Bull	z_2^2	0	0	0	0	0	0	0	0	0	0	72	0	0	2208	0	181	0	59
Bear	z_3^2	0	0	0	0	0	0	0	27	0	15	0	34	831	0	846	0	16	0
Bear	z_4^2	58	0	158	0	2155	0	0	22	0	0	0	0	0	1	0	0	0	0

Now, look how at how the features are classified: Mention plot number on tayal.

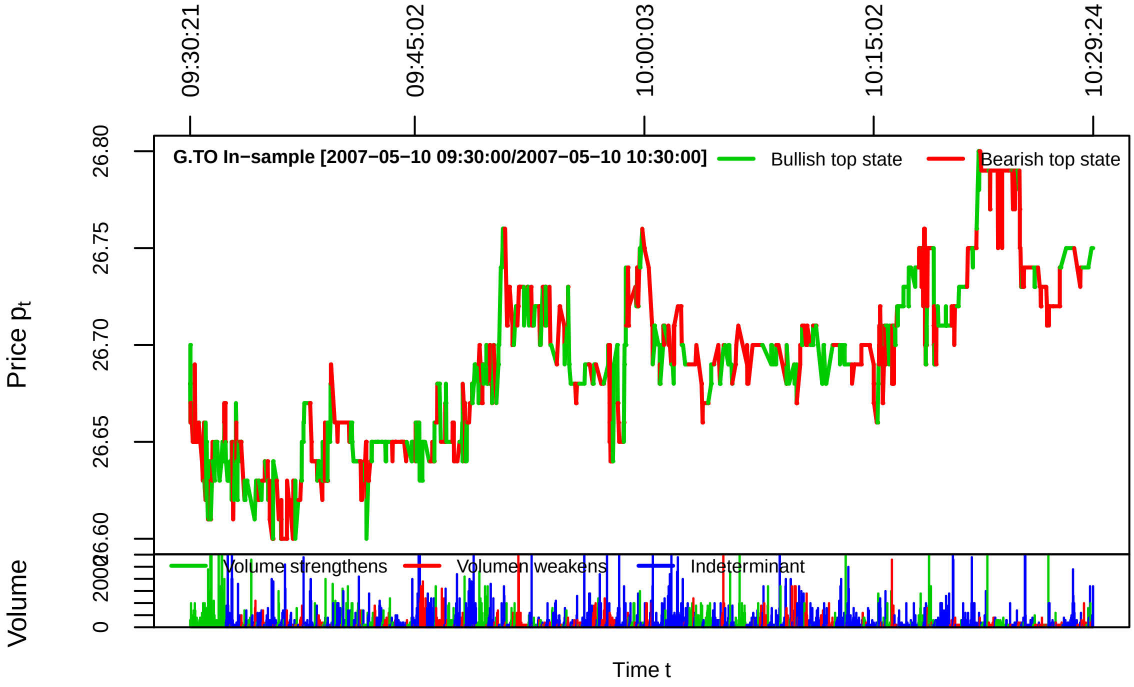

Figure 7: Tick by tick sequence of trades, classified as belonging to bear or bull states, from TSE:G 2007-05-10 09:30:00/2007-05-10 10:30:00.

...

3.6.6 In-sample summary

...

3.6.7 Forecast

...

3.6.7.1 Walking forward forecasts

...

3.6.8 Trading strategy

...

3.7 Discussion

3.7.1 The statistical model

...

3.7.2 The financial application

...

---

4 References

Fine, Shai, and Yoram Singer. 1998. "The Hierarchical Hidden Markov Model: Analysis and Applications."

Gelman, Andrew, and Donald B Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science*. JSTOR, 457–72.

Ord, Tim. 2008. "The Secret Science of Price and Volume." *Master Traders: Strategies for Superior Returns from Today's Top Traders*. Wiley Online Library, 87–105.

Sandoval, Javier, and Germán Hernández. 2015. "Computational Visual Analysis of the Order Book Dynamics for Creating High-Frequency Foreign Exchange Trading Strategies." *Procedia Computer Science* 51. Elsevier: 1593–1602.

Tayal, Aditya. 2009. "Regime Switching and Technical Trading with Dynamic Bayesian Networks in High-Frequency Stock Markets." Master's thesis, University of Waterloo.

Wisebourt, Shaul Sergey. 2011. "Hierarchical Hidden Markov Model of High-Frequency Market Regimes Using Trade Price and Limit Order Book Information." Master's thesis, University of Waterloo.

---

5 Appendix

```
## Warning in cbind(row.names, mat): number of rows of result is not a
## multiple of vector length (arg 1)
```

row.names	mean	se_mean	sd	10%	50%	90%	n_eff	Rhat
π_1	0.5067	0.0178	0.282	0.0956	0.5162	0.8931	250	0.9973
$1 - \pi_1$	0	0	0	0	0	0	250	NaN
a_{11}	0.4933	0.0178	0.282	0.1069	0.4838	0.9044	250	0.9973
a_{12}	0	0	0	0	0	0	250	NaN
$1 - a_{11} - a_{12}$	0	0	0	0	0	0	250	NaN
a_{22}	0.4559	0.0099	0.1089	0.3085	0.47	0.579	119.7043	1.0023
a_{31}	0.5441	0.0099	0.1089	0.421	0.53	0.6915	119.7043	1.0023
a_{33}	0	0	0	0	0	0	250	NaN
$1 - a_{31} - a_{33}$	1	0	0	1	1	1	250	NaN
a_{44}	0	0	0	0	0	0	250	NaN
π_1	0	0	0	0	0	0	250	NaN
$1 - \pi_1$	0	0	0	0	0	0	250	NaN
a_{11}	0.0914	0.0044	0.0623	0.0171	0.0836	0.1717	198.8658	0.9998
a_{12}	0	0	0	0	0	0	250	NaN

row.names	mean	se_mean	sd	10%	50%	90%	n_eff	Rhat
$1 - a_{11} - a_{12}$	0	0	0	0	0	0	250	NaN
a_{22}	0.9086	0.0044	0.0623	0.8283	0.9164	0.9829	198.8658	0.9998
a_{31}	0	0	0	0	0	0	250	NaN
a_{33}	0	0	0	0	0	0	250	NaN
$1 - a_{31} - a_{33}$	1	0	0	1	1	1	250	NaN
a_{44}	0	0	0	0	0	0	250	NaN
π_1	0.0073	6e-04	0.0057	7e-04	0.0064	0.0152	99.4159	1.0095
$1 - \pi_1$	0.0196	3e-04	0.0047	0.0145	0.0188	0.0263	250	0.997
a_{11}	0.0072	8e-04	0.0081	6e-04	0.0047	0.017	102.0471	1.0182
a_{12}	0.3419	0.0017	0.0208	0.314	0.3436	0.3665	150.0721	1.0051
$1 - a_{11} - a_{12}$	0.2231	0.002	0.0262	0.1873	0.2253	0.2561	171.5795	1.0062
a_{22}	0.3531	0.0011	0.0174	0.3305	0.3549	0.3719	250	0.9962
a_{31}	0.0264	3e-04	0.0053	0.0205	0.0263	0.0333	250	0.9963
a_{33}	0.0052	4e-04	0.0043	6e-04	0.0042	0.0109	144.0005	0.9979
$1 - a_{31} - a_{33}$	0.0161	3e-04	0.0046	0.0109	0.0155	0.0226	250	0.9997
a_{44}	0.0017	2e-04	0.0018	2e-04	0.0011	0.004	66.4428	1.0057
π_1	0.0226	3e-04	0.0053	0.0162	0.0223	0.0297	250	1.0017
$1 - \pi_1$	0.0017	1e-04	0.0017	1e-04	0.0012	0.0039	250	0.9993
a_{11}	0.0484	0.0033	0.0333	0.0062	0.046	0.0918	102.2995	0.9976
a_{12}	0.8008	0.0042	0.0403	0.7497	0.8026	0.8528	90.8209	0.9961
$1 - a_{11} - a_{12}$	0.0236	0.0012	0.0174	0.005	0.0203	0.0469	194.2454	0.997
a_{22}	0.0796	5e-04	0.008	0.0698	0.0798	0.0891	250	0.9962
a_{31}	0.0011	1e-04	0.001	1e-04	9e-04	0.0023	250	0.9961
a_{33}	0.0205	3e-04	0.0045	0.015	0.0204	0.0262	250	0.9965
$1 - a_{31} - a_{33}$	0.0105	5e-04	0.0038	0.0062	0.0106	0.0153	52.9642	0.9989
a_{44}	0.0023	1e-04	0.0022	3e-04	0.0018	0.0051	250	0.996
π_1	0.0251	3e-04	0.0052	0.0187	0.0251	0.0314	250	0.9964
$1 - \pi_1$	0.3554	0.003	0.0313	0.3159	0.3549	0.3936	111.4928	0.9961
a_{11}	0.1986	0.0036	0.0363	0.1525	0.1982	0.2421	99.991	0.9961
a_{12}	0.3884	0.0013	0.0198	0.3645	0.3886	0.4154	227.1397	0.9964
$1 - a_{11} - a_{12}$	0.0023	1e-04	0.0022	4e-04	0.0017	0.0054	250	0.996
a_{22}	0.0155	2e-04	0.0038	0.0111	0.015	0.0205	250	0.996
a_{31}	0.0018	1e-04	0.0019	2e-04	0.0013	0.0043	250	0.9962
a_{33}	0.0165	8e-04	0.0066	0.007	0.017	0.0251	75.1274	1.0015
$1 - a_{31} - a_{33}$	9e-04	1e-04	9e-04	1e-04	6e-04	0.002	250	0.9983
a_{44}	0.0558	0.001	0.011	0.0419	0.0574	0.0677	124.3395	1.0162
π_1	0.0212	0.0012	0.0158	0.0044	0.0185	0.0452	178.8572	1.0063
$1 - \pi_1$	0.8793	0.0023	0.0237	0.8476	0.8803	0.9057	110.8629	1.0212
a_{11}	0.0132	0.001	0.0122	0.0017	0.0088	0.0314	151.8973	1.0038
a_{12}	0.0019	1e-04	0.0017	3e-04	0.0014	0.0039	250	0.9966
$1 - a_{11} - a_{12}$	0.0098	5e-04	0.0052	0.0028	0.0098	0.0168	110.7242	0.9962
a_{22}	0.0014	1e-04	0.0014	1e-04	9e-04	0.0032	250	0.9991

6 Original Computing Environment