Commodity price dynamics through time scales

Author: Gilbert MBARA
Supervisor: Prof. Ryszard KOKOSZCZYNSKI

A thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the
Faculty of Economic Sciences
University of Warsaw

September 20, 2019
Declaration

I, Gilbert MBARA, declare that this thesis titled, “Commodity price dynamics through time scales” and the work presented in it are my own. I confirm that

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: 

Date: 
“A market-maker knows the price of everything and the value of nothing”

After Oscar Wilde
Summary

This dissertation develops models of commodity price dynamics based on the information flow that arises from trading in forward contracts. The models are presented in three distinct but related papers, each using data recorded at different observation time scales: from tick level price changes within a single trading session to aggregate movements over decades. For the micro-level analysis, the data considered are recorded at a transaction-by-transaction frequency within a single trading session at an exchange in a day. When the observation time scale is adjusted to the daily frequency, the time period of analysis is extended to data covering a year. Finally at the low-frequency monthly time scale, the analysis extends to cover time series spanning a period of 40 years. While a single time scale of analysis is used in each paper, the framework allows for easy transitioning to a different time scale.

Using data at the daily frequency, the first paper presents a parametric approach to estimating the underlying components of observed transaction prices of a commodity futures contract. Simple dynamic linear models with switching are used to estimate the efficient price process and effective bid–ask spreads which are treated as unobserved components in a state–space system. In simulation studies and empirical applications using daily commodity futures prices, it is demonstrated that the models deliver reliable inference on transaction costs and the order flow process over the trading day even in the absence of high-frequency transactions data.

Moving to a finer time scale, the second paper postulates a price updating rule for agents trading in a high frequency futures market. A theory of history dependent price formation where agents use the observed sequence of trade signs to make quote adjustments through a trading session is developed. The theory makes falsifiable empirical predictions about the behavior of prices that are tested using high–frequency transactions data from a commodity exchange. Price impact, the signed effect of a transaction on future quotations, and market volatility are shown to be a function of the observed sequence of trade signs and the effective measures of transaction costs.

Changing time scales from high– to low–frequency, the third paper models the long term behavior of commodity prices. A new double–mixture autoregressive econometric model for time series subject to abrupt changes is proposed and estimated. The model uses two independent Markov chains, one for price growth and another for its volatility, to describe the evolution of commodity price time series over the last 40 years. Unlike models currently found in the literature, the double mixture autoregression presented in the paper allows the persistence of price time series to change within volatility regimes, independent of whether prices are rising or falling. The model effectively identifies the
boom–bust characteristics of commodity prices observed over long periods of time.

All the data used in the analysis derive from transactions prices recorded at various commodity exchanges. The first paper uses daily summaries of trading activity – open, close, min and max prices – of selected futures contracts traded at two preeminent commodity exchanges: the Intercontinental Exchange and the Chicago Mercantile Exchange. The second paper uses transaction-by-transaction level price data for two of the most liquid "physically delivered" commodity futures contracts traded at the Tokyo Commodities Exchange: Gold and Platinum. Finally the third paper uses “spot price” data, representing the nearest maturity futures contract, from the International Monetary Fund database of commodity prices which are aggregated to a monthly frequency.
Acknowledgments

I would like to sincerely thank my supervisor Prof. Ryszard Kokoszczynski for his guidance and patience as I slowly progressed towards completing this work. Most of the empirical analysis undertaken in this dissertation are versions of hidden Markov models, to which I was first introduced by Michal Pakos. I will always be grateful to him for piquing my interest in financial economics and mathematics. Finally, I would not have become an economist, but for Prof. Joanna Tyrowicz who twice made it possible for me to live in Warsaw and expand my horizons.
## Contents

Declaration .................................................................................. iii  
Summary ....................................................................................... vii  
Acknowledgements ........................................................................ ix  
List of Tables ................................................................................ xiii  
List of Figures ............................................................................... xv  
Acronyms ..................................................................................... xvii  
Dedication ..................................................................................... xix  

### Introduction

1. **Chapter 1: New Open to Old Close: Signs and Spreads in Daily Prices**
   1.1 Introduction ........................................................................ 5  
   1.2 Dynamics of Transactions Prices ...................................... 7  
      1.2.1 Assumptions ................................................................. 8  
      1.2.2 State Space Representations and Estimation ............... 12  
   1.3 Alternative Spread Estimators ......................................... 15  
      1.3.1 The Gibbs Sampling Estimator .................................... 16  
      1.3.2 Low Frequency Estimators ......................................... 16  
   1.4 Simulation Study ............................................................... 18  
      1.4.1 High Frequency .......................................................... 18  
      1.4.2 Low Frequency .......................................................... 21  
   1.5 Empirical Applications .................................................... 23  
      1.5.1 Data Description ........................................................ 23  
      1.5.2 Estimation Results ..................................................... 24  
      1.5.3 Comparison to Daily High–Low (HL) ......................... 26  
   1.6 Conclusion ......................................................................... 27  

### Chapter 2: Price impact as reaction to order flow imbalance

   2.1 Introduction ......................................................................... 29  
   2.2 Transactions history dependent prices formation ............... 31  
      2.2.1 History dependent pricing ......................................... 32  
      2.2.2 Properties of the Price Process under Market Impact .... 34  
      2.2.3 Empirical Predictions ............................................... 36  
      2.2.4 Relation to Literature ............................................... 38
2.3 Calibration ............................................................................................................ 39
   2.3.1 State Space Representations and Estimation ................................................. 39
   2.3.2 Data Description ......................................................................................... 41
   2.3.3 Estimation Results .................................................................................... 43
2.4 Empirical Tests ..................................................................................................... 45
2.5 Conclusion ........................................................................................................... 49

Chapter 3: A Double Mixture Autoregressive Model of Commodity Prices 51
   3.1 Introduction .................................................................................................... 52
   3.2 Review of Markov Switching Models ............................................................... 53
      3.2.1 State Dependent Volatility and Persistence ............................................. 54
      3.2.2 Finite Sample Biases ............................................................................. 55
   3.3 A Double Mixture Autoregressive Model ....................................................... 56
      3.3.1 Likelihood function .............................................................................. 57
      3.3.2 Filtering .................................................................................................. 58
      3.3.3 Two Stage Estimation ............................................................................ 59
   3.4 Empirical Results ............................................................................................ 59
      3.4.1 First Stage Estimates: \( \mu \) .................................................................. 60
      3.4.2 Second Stage Estimates: \( \sigma \) .............................................................. 61
      3.4.3 State Inference Using 2nd Stage Estimates ............................................ 62
   3.5 Conclusion ....................................................................................................... 69

Appendix ..................................................................................................................... 71
   A Spread Estimates with Autocorrelated Trades ................................................. 71
   B Moments of a Two-State Markov Switching Model ......................................... 73

Bibliography ............................................................................................................... 76
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Parameter Estimates for High Frequency Data</td>
<td>20</td>
</tr>
<tr>
<td>Table 1.2</td>
<td>Distribution of High Frequency Spread Estimates</td>
<td>20</td>
</tr>
<tr>
<td>Table 1.3</td>
<td>Parameter Estimates for Low Frequency Data</td>
<td>22</td>
</tr>
<tr>
<td>Table 1.4</td>
<td>Descriptive Statistics of Sample Prices</td>
<td>25</td>
</tr>
<tr>
<td>Table 1.5</td>
<td>Parameter Estimates: Commodity Prices</td>
<td>26</td>
</tr>
<tr>
<td>Table 2.1</td>
<td>Summary Statistics of Prices over 5 minute Intervals</td>
<td>43</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Parameter Estimates of Switching Dynamic Linear Model</td>
<td>44</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Regression Tests of Relation between Lags-1 to $k$ and Lags-$k$ to $k+1$</td>
<td>46</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>First Stage Estimates</td>
<td>62</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Second Stage Estimates</td>
<td>66</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Unconditional Moments</td>
<td>66</td>
</tr>
<tr>
<td>Table A.1</td>
<td>Probability, Value, and Change of Trade Direction</td>
<td>71</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 1.1</td>
<td>Simulated Data</td>
<td>19</td>
</tr>
<tr>
<td>Fig. 1.2</td>
<td>Distribution of High Freq. Spread Estimates</td>
<td>21</td>
</tr>
<tr>
<td>Fig. 1.3</td>
<td>Labeling Accuracy: High Frequency</td>
<td>21</td>
</tr>
<tr>
<td>Fig. 1.4</td>
<td>Daily Simulated Price Series</td>
<td>22</td>
</tr>
<tr>
<td>Fig. 1.5</td>
<td>Labeling Accuracy: Low Frequency</td>
<td>23</td>
</tr>
<tr>
<td>Fig. 1.6</td>
<td>Estimated Underlying Price Process at Open: ( \hat{m}_{t,1} )</td>
<td>23</td>
</tr>
<tr>
<td>Fig. 1.7</td>
<td>Open, Close, Min and Max Futures Prices (3 Months)</td>
<td>24</td>
</tr>
<tr>
<td>Fig. 1.8</td>
<td>Log–Half Spread Estimates: High–Low, Close–High–Low and HMM</td>
<td>27</td>
</tr>
<tr>
<td>Fig. 2.1</td>
<td>Schematic Representation of a Limit Order Book</td>
<td>32</td>
</tr>
<tr>
<td>Fig. 2.2</td>
<td>Open, Close, Min and Max Transaction Prices, 24th April 2019</td>
<td>42</td>
</tr>
<tr>
<td>Fig. 2.3</td>
<td>Distribution of Returns and Traded Volumes</td>
<td>43</td>
</tr>
<tr>
<td>Fig. 2.4</td>
<td>Time Variation of Spreads</td>
<td>45</td>
</tr>
<tr>
<td>Fig. 2.5</td>
<td>Lag-1 to ( k ) Response Functions</td>
<td>47</td>
</tr>
<tr>
<td>Fig. 2.6</td>
<td>Lag-( k ) to ( k+1 ) Response Functions</td>
<td>47</td>
</tr>
<tr>
<td>Fig. 2.7</td>
<td>Volatility and the Spread</td>
<td>48</td>
</tr>
<tr>
<td>Fig. 3.1</td>
<td>log Prices</td>
<td>60</td>
</tr>
<tr>
<td>Fig. 3.2</td>
<td>First Stage Smoothened States Inference: Crude Oil</td>
<td>63</td>
</tr>
<tr>
<td>Fig. 3.3</td>
<td>First Stage Smoothened States Inference: Aluminum</td>
<td>64</td>
</tr>
<tr>
<td>Fig. 3.4</td>
<td>First Stage Smoothened States Inference: Rubber</td>
<td>65</td>
</tr>
<tr>
<td>Fig. 3.5</td>
<td>Smoothed Second Stage Bust State Inference</td>
<td>67</td>
</tr>
<tr>
<td>Fig. 3.6</td>
<td>Smoothed Second Stage Volatility State Inference</td>
<td>68</td>
</tr>
</tbody>
</table>
# Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>Aluminum</td>
</tr>
<tr>
<td>BVC</td>
<td>Book Volume Classification</td>
</tr>
<tr>
<td>CHL</td>
<td>Abdi-Ranaldo Close-High-Low Spread Estimator</td>
</tr>
<tr>
<td>CL</td>
<td>Crude Oil</td>
</tr>
<tr>
<td>CME</td>
<td>Chicago Mechantile Exchange</td>
</tr>
<tr>
<td>HL</td>
<td>Corwin Schultz High-Low Spread Estimator</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Models</td>
</tr>
<tr>
<td>ICE</td>
<td>International Coffee Exchange</td>
</tr>
<tr>
<td>JST</td>
<td>Japan Standard Time</td>
</tr>
<tr>
<td>KC/KT</td>
<td>Coffee Ticker Symbol</td>
</tr>
<tr>
<td>LOB</td>
<td>Limit Order Book</td>
</tr>
<tr>
<td>LR</td>
<td>Lee-Ready Algorithm</td>
</tr>
<tr>
<td>MLE</td>
<td>Maximum Likelihood Estimation</td>
</tr>
<tr>
<td>MRR</td>
<td>Madhavan, Richardson and Roomans (1997) model</td>
</tr>
<tr>
<td>MS</td>
<td>Markov Switching</td>
</tr>
<tr>
<td>NBER</td>
<td>National Bureau of Economic Research (United States)</td>
</tr>
<tr>
<td>NG</td>
<td>Natural Gas Ticker Symbol</td>
</tr>
<tr>
<td>OLS</td>
<td>Ordinary Least Squares</td>
</tr>
<tr>
<td>PA</td>
<td>Palladium Ticker Symbol</td>
</tr>
<tr>
<td>QR</td>
<td>Quote Rule</td>
</tr>
<tr>
<td>RB</td>
<td>Rubber</td>
</tr>
<tr>
<td>TCE/JAU</td>
<td>Gold Ticker/Trading Symbol at TOCOM</td>
</tr>
<tr>
<td>TCE/JPL</td>
<td>Platinum Ticker/Trading Symbol at TOCOM</td>
</tr>
<tr>
<td>TOCOM</td>
<td>Tokyo Commodity Exchange</td>
</tr>
<tr>
<td>TR</td>
<td>Tick Rule</td>
</tr>
</tbody>
</table>
Dedication

To my parents,

Charles Juma & Elizabeth Siangla.
Introduction

Commodity markets are distinct from other product markets due to the existence of forward sales and futures contracts. Forward selling and the trading of a commodity derivative implies prices are subject to the influence of economic agents who are not directly engaged in consumption or production of the commodity. As a result, even when the forces of supply and demand are in equilibrium, prices may still move and vary purely due to activities of agents operating in the futures markets. Inspired by this observation, the dissertation provides a new analysis of the role of futures markets trading on the dynamics of commodity prices over different time scales.

Throughout the dissertation, the underlying economy can be conceived of as populated by a commodity producing firm with access to a stochastic production technology that yields new output of the commodity every time period. The firm also has access to a storage technology which it can use to hold inventory. The firm’s sales are made either in a spot market for cash or can be sold ahead of production for future delivery using a forward contract specifying the price and date of delivery to the holder or buyer. When such forward contracts are traded or exchanged in a centralized market, they become futures contracts.

Stochastic production and consumption of the commodity implies that the firm faces a risk of losses from volatile prices. The firm would therefore like to sell forward as much of its output as it can. Those who buy the firm’s contract take on the risk of changing prices (are subject to loss) and demand a risk-premium as compensation for taking over the firm’s risk. In commodity markets, this risk premium is measured either as the basis, the contemporaneous difference between the current spot price and the forward price or the expected return, the difference between the expected future spot price and the forward price (Yang, 2013). These risk premiums depend on transaction costs incurred when trading futures in a commodity exchange (Hasbrouck, 2009) and will have an effect on the firms investments in physical production of the commodity – as it reflects the cost of hedging.

Given that the risk premiums are a function of the transaction costs, quantifying the size of these costs has become an important endeavor in understanding of price dynamics. The first paper takes on this task by developing parametric models that can be used to measure liquidity costs using exchange traded futures transactions prices only. Simple dynamic linear regressions with switching are used in this task. The models treat underlying price processes and liquidity costs as unobserved components in state space systems with trade direction indicators of buyer and seller initiated transactions being the outcomes of hidden Markov processes. Simulation studies show that the model provides accurate effective transaction cost estimates and beats the tick-rule method of signing trades using
prices.

Having developed a way to accurately measure the liquidity costs, focus turns to what is driving price changes observed at a high frequency tick-by-tick level. The second paper presents a new theory of history dependent price setting in limit order book market for commodity futures. In traditional financial markets theory, the price discovery process is a form of tâtonnement; informed agents trading against liquidity providers or market markers slowly reveals private information which is incorporated into prices. The market-markers adjust their quotations to reflect the information revealed by the informed agents transactions until a new equilibrium is attained. However, when trading contracts of physically delivered commodities, the transactions are directly informative of expected future supply and demand since they reflect production and consumption intentions. Transactions therefore have price impact: a buyer initiated trade tends to push prices upwards with the opposite effect following a seller initiated trade. The history dependent framework takes this hypothesis to the data and shows that agents trading in a limit order book market for commodity futures adjust their prices in response to order flow – the sequence of trade originator signs.

Over long time periods, commodity price time series exhibit boom–bust cycles that may be accompanied by periods of either high or low volatility. One way to model time series subject to such boom and bust cycles is the hidden Markov or regime switching model popularized in economics by Hamilton (1990). The standard regime switching model assumes that the growth and volatility phases of a time series coincide and that autoregressive lag lengths are similar across regimes. This assumption results into biases in estimates of unconditional variances across different regimes. To overcome these problems, the third paper presents a new “Double Mixture Autoregressive” model for time series subject to potentially independent changes in level and volatility. This model allows for the autocorrelation structure of the data generating process to vary across variance regimes. By accounting for the change in the lag length of time series across the different volatility periods, more precise estimates of the unconditional moments are obtained. The model is applied to set of industrial commodity prices and is shown to accurately represent the boom–bust cycles and volatility switches that characterize the time series.

This dissertation is divided into three related chapters/papers. The first chapter/paper, “New Open to Old Close: Signs and Spreads in Daily Prices” presents state space models that can be used to obtain accurate measures of transaction costs using daily summaries of trading activity: open, close, max and min prices. The second chapter/paper, “Price impact as reaction to order flow imbalance”, develops and successfully tests a theory of history dependent price formation in a limit order book market of commodity futures. Finally, the third chapter/paper, “A Double Mixture Autoregressive Model of Commodity Prices”, presents a new type of non-linear econometric model that captures the boom–bust cycles and volatility switches that characterize the long term behavior of commodity price time series.

The three papers use different time scales of analysis. The first uses commodity price
series at a daily frequency for periods of up to one year, approximately 252 days of market open to close futures prices from a commodity exchange. The second paper uses high-frequency trade-by-trade or tick-level data from a continuous trading session in a single day. Finally the third paper uses long time series spanning decades. It is important to look at data from the microscopic (tick-level) to the macroscopic (decades) time scale in order to obtain a holistic view of the behavior of prices. Over very short time periods, the tick size is the smallest movement over any two prices and price changes can be viewed as random walks over a grid, with jumps occurring at arbitrary times (Curato and Lillo, 2014). This calls for the modeling of the microstructural features of the data such as the bid–ask spread; usually equal to half a tick for highly liquid assets, and the sequence of buy–sell orders which may predict the direction of short term price movements. At a coarser time-scale of months, quarters or years, the microstructural issues can be dispensed with and a more traditional time series approach used to describe price dynamics. While I separate the analysis based on the observation time scales, multi frequency models of price dynamics are possible, albeit with a more complex structure to capture the volatility components at play over every time scale (Calvet and Fisher, 2001, 2004).
CHAPTER 1

New Open to Old Close: Signs and Spreads in Daily Prices

Abstract

I present simple dynamic linear models with switching to estimate transaction costs in financial markets. The models treat underlying price processes and bid–ask spreads as unobserved components in state space systems with indicators of buyer or seller initiated transactions being the outcome of a hidden Markov process. In simulation studies, I show that the model provides accurate spread estimates and beats the tick-rule method of signing trades using prices: 82% versus 66% labeling accuracy of transactions from small samples. The model easily transitions to a low-frequency data implementation without loss of precision in parameter estimates. In empirical applications using daily commodity futures prices, I show that the model can deliver reliable inference on transaction costs and the order flow process over the trading day even in the absence of high-frequency intraday data.

Keywords: transaction costs; bid–ask spread; signing trades; tick rule; market microstructure; price dynamics.

1.1 Introduction

One goal of financial markets is to offer investors a fast and cheap way to execute their portfolio transactions. A measure of how well the market achieves this purpose is the transaction costs associated with making a trade. A commonly used proxy for these costs is the effective bid–ask spread: the difference between prices of buyer and seller initiated transactions. In the classical analysis of trading in financial markets, the spread exists as a means of compensating a “market maker” or another market participant for providing liquidity services. Transaction costs, which include the bid-ask spread, can reduce trading in an asset and result into a large “illiquidity discount” (Lo, Mamaysky and Wang, 2004). Liquidity is the cost of immediacy and the bid-ask spread is a non-negligible component of this cost (Plerou, Gopikrishnan and Stanley, 2005).

1The spread may be decomposed into three separate cost components: order-processing, inventory and adverse selection costs (Glosten and Harris, 1988; Biais, Glosten and Spatt, 2005).
1.1. Introduction

Estimating the spread requires knowledge of the initiator side for any observed transaction price. This information is usually unavailable or unobservable and has to be imputed from trade classification algorithms or rules (Boehmer, Grammig and Theissen, 2007). The tick rule (TR) assigns a trade to be a buy if the price was an up-tick relative to the previous trade and to be a sell if it was a down-tick, while the quote rule (QR) classifies a trade as buyer (seller) initiated if the trade occurs at a price above (below) the midpoint; an average of the best–bid and best–ask prices. A popular method that combines the two rules to sign trades is the Lee–Ready (LR) algorithm of Lee and Ready (1991).\(^2\) There is disagreement on the accuracy of these classifiers, with success dependent on context. For instance, Blais and Protter (2012) have found that the LR algorithm can only accurately classify around 61% of trades for a highly liquid security.\(^3\) Misclassification has consequences for transaction cost estimates: Theissen (2001) find that estimates based on the tick rule grossly overstate the spread, Werner (2003) shows that the order composition (buys/sells) of trades affects traditional measures of spreads and Boehmer et al. (2007) show that estimates of the probability of informed trading (PIN)\(^4\) are downward biased in the presence of misclassified trades. In a nutshell, accurate measures of transaction costs depend on how well one can classify the observed transaction prices.

The LR and QR algorithms require snapshots of the limit order book (LOB)\(^5\) throughout the trading day and are inapplicable on data sets that only record transaction prices. However, when LOB data are available, the computing infrastructure required to handle the data is non–trivial for any time series spanning periods as long as a year. For instance, to store 5 years of LOB data for 100 of the most liquid futures contracts with around 3 billion trades, Wu, Bethel, Gu, Leinweber and Rübel (2013) used about 140 GB of text files and required parallelization to achieve fast computations. The computational burden remains severe even when using transaction prices only. For high frequency trade–by–trade data, a common parametric approach for estimating spreads is the Gibbs sampling estimator of Hasbrouck (2004, see Sec. 1.3.1). Wel, Menkveld and Sarkar (2009) compared a Markov switching state space maximum likelihood method (similar to the one proposed here) to the Gibbs sampler and found that for their data, the sampler would take 40 hours to estimate for a given model specification in contrast to the likelihood method which took 110 minutes. Methods that rely on high frequency LOB data for measuring time varying liquidity consequently require expensive computing infrastructure and/or data compression techniques such as the clustering approach suggested by Buza, Nagy and Nanopoulos (2014).

To address these challenges, I propose a parametric method to classify trades and estimate time-varying bid–ask spreads using only transaction prices. I assume that the ob-

\(^2\)The LR algorithm first applies the QR to all trades and followed by the TR on trades at the midpoint.
\(^3\)See for example Aktas and Kryzanowski (2014); Perlin, Brooks and Dufour (2014); Chakrabarty, Pascual and Shkiklo (2015) for an evaluation of these algorithms.
\(^4\)Trading against an informed agent is the adverse selection cost faced by a typical market marker.
\(^5\)The LOB collects all the buying and selling interest in a particular security and presents an aggregation of these orders to every market participant.
served trade price consists of three unobserved components: an "efficient" price process, a bid–ask spread and a trade initiator indicator. A trade occurs at the bid if it is seller initiated and at the ask if it is buyer initiated, with the spread being the difference between the two prices. The data do not include a label on who the initiator side of the trade was and the transaction label has to be determined jointly with the value of the spread. I model the trade initiator (or direction) indicator as the outcome of a hidden first order two–state Markov process, while the spread is time varying with a constant mean and non–zero variance. Together, the efficient price process, the trade direction indicator and the time varying spread form a dynamic linear model with switching which can be estimated for high–frequency trade–by–trade or low–frequency daily data.

In a simulation study, I show that the estimation recovers parameters of the observed transactions price processes: the underlying efficient price, spreads and trade initiator indicators. For the effective spread and parameters related to the efficient price process, the model performance matches the Gibbs sampling estimator of Hasbrouck (2004). In classification, the model outperforms the TR algorithm and produces superior labeling accuracy rates in comparison to those reported in the literature. There is currently active development of low frequency estimators that use daily summaries of trading activity: open, close, high and low prices to approximate transaction costs. I implement a daily version of the simulation study by discarding the intermediate trades that occur during a typical trading day and re-estimating the model using daily open to close and high–low prices only. The model performance is similar to the high–frequency case in initiator labeling while giving non-negative spread estimates that do not require any adjustments as the High–Low and Close–High–Low transactions costs estimators of Corwin and Schultz (2012) and Abdi and Ranaldo (2017).

The remainder of the paper is organized as follows. Sec. 1.2 develops the model assumptions and their implications for the dynamics of prices. The initial model description assumes high–frequency trade by trade data are observed, but simple modifications of the specification easily transition the estimation to data based on summaries of the daily trading activity. Sec. 1.3 describes alternative measures or estimators of the effective spread. Sec. 1.4 simulates trade–by–trade transactions prices to demonstrate the model’s ability to recover trade initiator labels and bid–ask spreads in both high– and low–frequency environments. Sec. 1.5 performs an empirical application using daily data that records only summaries of market activity for three commodity futures contracts: coffee, natural gas and palladium. Sec. 1.6 concludes.

1.2 Dynamics of Transactions Prices

Roll (1984) suggested a simple model of security prices in the presence of transaction costs. In his model, the observed trade price is assumed to consist of an efficient price

---

6These assumptions do not affect the robustness of the results as is shown by comparing to alternative estimators discussed in Sec. 1.3.
1.2. Dynamics of Transactions Prices

component that follows a geometric Brownian motion plus a transaction cost component whose value depends on whether the transaction is buyer or seller initiated. Concretely, let $p_t$ denote the log transaction price at date $t$ of a given security and let $m_t$ denote the log efficient price that would prevail in a frictionless market without any transaction costs. The dynamics of prices are:

$$p_t = m_t + s_t q_t; \quad m_t = m_{t-1} + u^m_t, \quad s_t = s$$

(1.1)

where $s_t = s$ is a constant bid-ask spread and $q_t$ is a trade direction indicator defined by:

$$q_t = \begin{cases} +1 & \text{if transaction is buyer initiated} \\ -1 & \text{if transaction is seller initiated} \end{cases}$$

with $u^m_t$ being a zero-mean disturbance uncorrelated with $q_t$. Roll motivates $s_t = s$, a constant, as one-half the quoted bid-ask spread, but since the model refers to transaction prices, $s$ can be viewed as the effective transaction cost (Hasbrouck, 2009). In equation (1.1), a buyer initiated transaction generates a price that exceeds the efficient price by $s$, while a seller initiated transaction has the opposite effect. I make the following assumptions:

1.2.1 Assumptions

Assumption 1: (Basic Roll). Observed transaction prices \( \{p_t\}_{t=1}^T \) are generated from equation (1.1) where \( \{u^m_t, q_t\}_{t=1}^\infty \) is a strictly stationary process with $u^m_t \sim N(0, \sigma^2_m)$.

Assumption 2: (Random Spread). The spread $s_t$ is a random variable that follows the process:

$$s_t = s + u^s_t, \quad \text{where } u^s_t \sim N\left(0, \sigma^2_s\right) \quad \text{with} \quad u^s_t \perp u^m_t.$$ 

(1.2)

Equation (1.2) allows the spread to vary around a constant mean $E[s_t] = s$ in response to changes in market conditions unrelated to the efficient price process (see e.g. Figure 1 of Amaya, Filbien, Okou and Roch, 2018). Representations of a time-varying spread have been made by Bandi, Pirino and Renò (2017) who write $s_t = \alpha (s + \gamma |p^e_t - m_t|)$ where $\alpha = \frac{s + s^2}{s^2 + \gamma^2 s^2}$, $p^e_t, m_t$ respectively being the efficient and midpoint prices which lead to time varying spreads while guaranteeing $E[s_t] = s$; and by Chen, Linton and Yi (2017) who don’t make any distributional assumptions about the process followed by $s_t$. For cross-sectional data, Hasbrouck (2009) assumed $s_{it} = \gamma_{0i} + \gamma_{1i} z_t$ for stocks indexed by $i$ with $z_t$ being an unobserved factor common to the effective costs across all firms.
**Assumption 3:** (Markov Trades). The trade direction indicator is the outcome of a first order Markov process defined by the transition matrix:

$$
P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \quad (1.3)
$$

where the entries $p_{jk}, j,k = 1,2$ are transition probabilities defined by $\text{Prob}(q_t = k|q_{t-1} = j,q_{t-2} = l,\ldots) = \text{Prob}(q_t = k|q_{t-1} = f) = p_{jk}$ for $q_t = \pm 1$. The assumption of the trade direction indicators following a Markov process has been made by Madhavan, Richardson and Roomans (1997, hereafter MRR) and Huang and Stoll (1998) among others.\(^7\) This assumption has several implications for the order book process which I now briefly discuss.

**Remark 1: Autocorrelation of Trades** The transition matrix $P$ implies the unconditional probabilities of being in states $q_t = \pm 1$ at time $t = 0$, $\pi_1 = \text{Prob}(q_t = +1)$ and $\pi_2 = \text{Prob}(q_t = -1)$, are given by:

$$
\pi_1 = \frac{P_{21}}{P_{12} + P_{21}}; \quad \pi_2 = \frac{P_{12}}{P_{12} + P_{21}}.
$$

In addition, the transition matrix $P$ has eigenvalues $\lambda = (\lambda_1, \lambda_2) = [1, 1 - P_{12} - P_{21}]$ with corresponding eigenvectors $v_1 = (1,1)'$ and $v_2 = (-P_{12}, P_{21})'$, so we can write $P = V \lambda V^{-1}$ where $D$ is a matrix with the eigenvalues of $P$ on its main diagonal and $V$ a matrix whose columns are given by the eigenvectors $[v_1, v_2]$. It follows that for $n \geq 0$,

$$
P^{(n)} = V \lambda^n V^{-1} = \begin{bmatrix} \pi_1 + \pi_2 \lambda_1^n & \pi_2 \left(1 - \lambda_2^n\right) \\ \pi_1 \left(1 - \lambda_2^n\right) & \pi_2 + \pi_1 \lambda_2^n \end{bmatrix} \quad (1.4)
$$

and $P^{(n)}$ is the probability of moving from one state to the next in $n$ steps. The mean and variance of $q_t$ are respectively given by: $q = \mathbb{E}[q_t] = \pi_1 - \pi_2$ and $\text{Var}(q_t) = \mathbb{E}[q_t^2] - \mathbb{E}[q_t]^2 = 4\pi_1\pi_2$. Conditioning on $q_t$, we can use the law of iterated expectations to compute $\mathbb{E}[q_{t+n}|q_t]$ as:

$$
\mathbb{E}(q_{t+n}|q_t) = \mathbb{E}_{q_t}\left[\mathbb{E}\left(q_{t+n}|q_t\right)\right] = \pi_1 \cdot \mathbb{E}(q_{t+n}|q_t = +1) - \pi_2 \cdot \mathbb{E}\left(q_{t+n}|q_t = -1\right)
$$

$$
= \pi_1^2 + \pi_2^2 - 2\pi_1\pi_2 \left(1 - 2\lambda_2^n\right),
$$

where $\mathbb{E}[q_{t+n}|q_t = +1]$ is given by:

$$
\mathbb{E}[q_{t+n}|q_t = +1] = 1 \cdot \text{Prob}\left[q_{t+n} = +1|q_t = +1\right] = 1 - \pi_1 + \pi_2 \lambda_1^n - \pi_2 \left(1 - \lambda_2^n\right)
$$

\(^7\)The MRR model further assumes that some trades can occur at the midpoint; such trades being neither buyer nor seller initiated. Midpoint trades can be easily accommodated by considering an extra value for the trade direction indicator: $q_t = (+1,0,-1)$ and defining a $(3 \times 3)$ transition matrix: $P = \begin{bmatrix} P_{11},P_{00},P_{12} - P_{00} \\ \frac{1}{2}(1 - P_{00}), -P_{00}, \frac{1}{2}(1 - P_{00}) \\ P_{21} - P_{00}, P_{00}, P_{22} \end{bmatrix}$, with $p_{00}$ being the probability of consecutive trades at the midpoint (see Ranaldo, 2008).
1.2. Dynamics of Transactions Prices

and \( E[q_{t+n}|q_t = -1] \) is similarly computed. We then have, for non-negative integers \( n \), the autocorrelation function \( \rho^{(n)} \):

\[
\rho^{(n)} = \frac{E[q_{t+n}q_t] - E[q_{t+n}]E[q_t]}{\text{Var}[q_t]} = \lambda_2^n = \left(1 - p_{12} - p_{21}\right)^n
\]

(1.5)

Equation (1.5) implies that if \( p_{11} = p_{22} = \frac{1}{2} \), then \( \pi_1 = \pi_2 = \frac{1}{2} \) and a buy or a sell is equally likely at any point in time as has been assumed in some studies (Hasbrouck, 2004). It would then follow that \( \rho^{(n)} = 0 \) at all leads and lags.

**Remark 2: Order Flow Process** The two state Markov process admits the idea, originally due to Hasbrouck (1988), that the unexpected rather than actual order flow affect the price process. We can compute:

\[
E[q_t|q_{t-1} = 1] = 1 \cdot \text{Prob}[q_t = +1|q_{t-1} = +1] - 1 \cdot \text{Prob}[q_t = -1|q_{t-1} = +1]
\]

\[
= (p_{11} - p_{12}) q_{t-1}
\]

\[
= q(1 - \rho) + \rho q_{t-1} = E[q_t|q_{t-1} = -1],
\]

(1.6)

where \( q = E[q_t] \) and \( \rho \) is defined by (1.5). The last equality in (1.6) means that we can write the conditional expectation as a regression function, we can write the trade direction indicator as an autoregressive process with a drift:

\[
q_t = q(1 - \rho) + \rho q_{t-1} + u_t^q, \quad E[u_t^q] = 0, \quad \text{Var}(u_t^q) = \frac{4\pi_1\pi_2}{1 - \rho^2}.
\]

(1.7)

The order flow surprise of the MRR model is then given by: \( q_t - E[q_t|q_{t-1}] = u_t^q \), which is an unpredictable component of trades. Under Assumption 3, only the unpredictable part of the order flow would have any impact on the underlying price process of the MRR model.

**Remark 3: Expanded State Space** Based on the transition matrix (1.3), we can now define a process for the price change equation with the hidden state vector \( \vec{q}_t = (q_t, q_{t-1}) \):

\[
\Delta p_t = \Delta m_t + s_t q_t - s_{t-1} q_{t-1},
\]

(1.8)

where the state vector \( \vec{q}_t \) takes on the values:

\[
\vec{q}_t = (q_t, q_{t-1}) = \left\{ \left[ +1, +1 \right], \left[ +1, -1 \right], \left[ -1, +1 \right], \left[ -1, -1 \right] \right\}
\]

(1.9)

In equation (1.8) the price change \( \Delta p_t \) can take on four values following the values of \( \vec{q}_t \) defined in equation (1.9). The hidden state \( \vec{q}_t \) is first order Markov with transition matrix
\[ \hat{P} \text{ whose entries are defined by: } \hat{p}_{jk} = \text{Prob}(\tilde{q}_t = k|\tilde{q}_{t-1} = j, \tilde{q}_{t-2} = l, \ldots) = \text{Prob}(\tilde{q}_t = k|\tilde{q}_{t-1} = j) \text{ for } j, k, l \in \{\tilde{q}_t\}: \]

\[ \hat{P} = \begin{pmatrix} p_{11} & 0 & p_{12} & 0 \\ p_{11} & 0 & p_{12} & 0 \\ 0 & p_{21} & 0 & p_{22} \\ 0 & p_{21} & 0 & p_{22} \end{pmatrix} \]  

(1.10)

where the \( p_{jk} \) are given by (1.3) and preserve the normalization \( \sum_{k=1}^{4} \hat{p}_{jk} = 1 \). The entries of this \( 4 \times 4 \) matrix follow directly from the relationship between \( q_t \) and \( q_{t-1} \) given in equation (1.9). The expansion of the state-space and its accompanying transition probability matrix follows directly from Assumption 3 on first order Markov property of trades. Similar expansions of the state space have been used at least since the work of Cox and Miller (1965).

**Remark 4: Low Frequency Transition**  For long time series spanning decades, it is usually difficult and/or impractical to obtain trade-by-trade data to estimate transaction costs. However, summaries of the trading day are generally available for many markets, sometimes for periods of more than 30 years. Theses summaries include the daily open, close, high and low prices as well as traded volumes. There is consequently an ongoing effort to develop “low frequency” estimators which rely solely on daily data. These include the High-Low (HL) and Close–High–Low (CHL) estimators of Corwin and Schultz (2012) and Abdi and Ranaldo (2017), respectively; which are described in more detail in Sec. 1.3.2.

To exploit the full range of information available from daily price time series, we can rewrite the price change equation (1.8) as:

\[ \Delta p_t = p_{t,1} - p_{t-1,n} = m_{t,1} - m_{t-1,n} + s_{t,1}q_{t,1} - s_{t-1,n}q_{t-1,n}, \]  

(1.11)

where \( p_{t,1} \) is the log open or first price on day/date \( t \) and \( p_{t-1,n} \) is the log close or last price on the previous day/date. The efficient price process \( m \) and the bid–ask spread \( s \) are similarly dated. The expanded state vector \( \tilde{q}_t \) takes on the values similar to equation (1.9):

\[ \tilde{q}_t = (q_{t,1}, q_{t-1,n}) = \left\{ [1, 1], [1, -1], [-1, 1], [-1, -1] \right\}. \]

However, the transition matrix now has to take into account that there are intermediate trades occurring within the day that are discarded in the specification. The transition matrix becomes:

\[ \hat{P} = \begin{pmatrix} p_{11} \cdot p_{11}^{(m)} & p_{21} \cdot p_{11}^{(m)} & p_{12} \cdot p_{11}^{(m)} & p_{22} \cdot p_{11}^{(m)} \\ p_{11} \cdot p_{12}^{(m)} & p_{21} \cdot p_{12}^{(m)} & p_{12} \cdot p_{12}^{(m)} & p_{22} \cdot p_{12}^{(m)} \\ p_{11} \cdot p_{21}^{(m)} & p_{21} \cdot p_{21}^{(m)} & p_{12} \cdot p_{21}^{(m)} & p_{22} \cdot p_{22}^{(m)} \\ p_{11} \cdot p_{22}^{(m)} & p_{21} \cdot p_{22}^{(m)} & p_{12} \cdot p_{22}^{(m)} & p_{22} \cdot p_{22}^{(m)} \end{pmatrix}. \]  

(1.12)
1.2. Dynamics of Transactions Prices

The probabilities $p_{jk}^{(n)}$, of moving from states $j$ to $k$ in $n$ steps are defined by the matrix $P^{(n)}$ of equation (1.4). Entries in the transition matrix (1.12) are obtained using the general product rule: $\text{Prob}(A,B|C,D) = \text{Prob}(B|C,D) \cdot \text{Prob}(A|B,C,D)$ for any events $A, B, C, D$ and the Markov property of $q_t$. For instance, the (row, column) = (3, 1) entry of $\hat{P}$ is defined by:

$$\text{Prob}[\tilde{q}_t = (1)|\tilde{q}_{t-1} = (3)] = \text{Prob}[q_{t,1} = +1, q_{t,1,n} = +1|q_{t-1,1} = -1, q_{t-2,n} = +1] = \text{Prob}[q_{t,1} = +1|q_{t-1,n} = +1] \cdot \text{Prob}[q_{t-1,1} = -1, q_{t-2,n} = +1]$$

where the last equality has used the first order Markov property of $q_t$ and the other entries are similarly derived. Note that $\hat{P}$ is a stochastic matrix as its rows add up to unity: $\Sigma_{k=1}^{4} \hat{P}_{jk} = 1, \forall k$.

1.2.2 State Space Representations and Estimation

**State Space Representations:** The process defined by equation (1.8) can be written in state space form as:

$$y_t = \Delta p_t = m_t - m_{t-1} + s_t q_t - s_{t-1} q_{t-1} = A_t x_t,$$

where $x_t = \left( m_t, m_{t-1}, s_t, s_{t-1} \right)'$ defines a (4 × 1) unobserved state vector that follows the first autoregressive process $x_t = \Phi x_{t-1} + \Upsilon + u_t$:

$$\begin{pmatrix} m_t \\ m_{t-1} \\ s_t \\ s_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} m_{t-1} \\ m_{t-2} \\ s_{t-1} \\ s_{t-2} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ s \\ 0 \end{pmatrix} + \begin{pmatrix} u_t^m \\ u_t^s \\ u_t \\ 0 \end{pmatrix} \quad (1.14)$$

and $A_t = \left[1, -1, q_t, -q_{t-1}\right]$ is a (1 × 4) observation or measurement matrix that takes the values:

$$A_t = \{ \tilde{A}_1, \tilde{A}_2, \tilde{A}_3, \tilde{A}_4 \} = \{ [1, -1, 1, -1], [1, -1, 1, 1], [1, -1, -1, -1], [1, -1, -1, 1] \} \quad (1.15)$$

The observation matrix $A_t$ is a first order Markov process with transition probability matrix defined by $\hat{P}$ given in equation (1.10). The error vector $u_t = (u_t^m, 0, u_t^s, 0)'$ is i.i.d $N(0, \Sigma_u)$ where $\Sigma_u$ is a (4 × 4) variance–covariance matrix with off–diagonal elements equal to zero and diagonal $(\sigma_m^2, 0, \sigma_s^2, 0)$. Similarly, the process defined by equation (1.11) can be written in state space form as

$$\Delta p_t = p_{t,1} - p_{t-1,n} = m_{t,1} - m_{t-1,n} + s_{t,1} q_{t,1} - s_{t-1,n} q_{t-1,n} = A_t x_t, \quad (1.16)$$
where the state $x_t$ is now defined by:

$$
\begin{pmatrix}
  m_{t,n} \\
  m_{t,1} \\
  m_{t-1,n} \\
  s_{t,1} \\
  s_{t-1,n}
\end{pmatrix} =
\begin{pmatrix}
  0 & 0 & 0 & 0 & 0 \\
  1 & 0 & 0 & 0 & 0 \\
  1 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 1 \\
  0 & 0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
  m_{t-1,n} \\
  m_{t-1,1} \\
  m_{t-2,n} \\
  s_{t-1,1} \\
  s_{t-2,n}
\end{pmatrix} +
\begin{pmatrix}
  0 \\
  0 \\
  0 \\
  s \\
  0
\end{pmatrix}
\begin{pmatrix}
  u^m_{t,1} \\
  0 \\
  u^s_{t,1}
\end{pmatrix}
$$

(1.17)

and $A_t = [0, 1, -1, q_{t(1)}, -q_{t-1(1)}]$ is a $(1 \times 5)$ observation or measurement matrix that takes the values similar to equation (1.15) with transition probability matrix $P$ given in equation (1.12). The entry $\hat{m}_{t,n}$ in equation (1.17) is a proxy for the end of day efficient price. Abdi and Ranaldo (2017) have argued that the mid-quote of day $t$, defined as the average of midranges over days $t$ and $t + 1$ is an unbiased proxy for $m_{t,n}$ (see Sec. 1.3.2 below).

Finally, the order flow process equation (1.7) allows us to dispense with the switching structure in the following way. The product $s_t q_t$ can be approximated using a first-order Taylor series expansion around $\mathbb{E}(s_t) = s$ and $\mathbb{E}(q_t) = q$: $s_t q_t \approx -sq + qs_t + sq_t$ with the difference $\Delta(s_t q_t) \approx s \cdot \Delta q_t + q \cdot \Delta s_t$. Using this approximation, equation (1.13) can be rewritten with the measurement, state and error:

$$
A_t = [1, -1, s, -s, q, -q], \quad x_t = \left(m_t, m_{t-1}, q_t, q_{t-1}, s_t, s_{t-1}\right)', \quad u_t = \left(u^m_t, 0, u^q_t, 0, u^s_t, 0\right)',
$$

(1.18)

respectively. This version of the model is simpler (more linear) even though it involves the estimation of one extra parameter in comparison to the switching regression. For the transition probabilities $p_{11}$ and $p_{22}$, we instead have to estimate $q, \rho$ and $\sigma_q$. While this linearized approximation converges faster in estimation, it does not give accurate inference on the trade direction/initiator indicator, even though such inference can still be made.\(^8\) I now discuss the estimation of this model.

**Maximum Likelihood Estimation**: I use maximum likelihood methods, even though other authors have used Bayesian techniques for state space models with switching (see e.g. Douc and Stoffer, 2014, Chapter 10). Estimation is facilitated by using the Hamilton filter (Hamilton, 1994). The parameters to be estimated are $\Theta_1 = \{s, \sigma^m, \sigma^q\}$ and $\Theta_2 = \{p_1, p_2\}$; the last two defining the transition matrix probabilities $p_{kk} = \frac{\exp{p_k}}{1+\exp{p_k}}, k = \{1, 2\}$. Letting $\epsilon_{t,\ell} = (y_t - \delta\ell x_t^{-1})$ denote the innovation from equation (1.13) when $A_t = \delta\ell, \ell = 1, 2, 3, 4$, then the log-likelihood is given by:

$$
\ln L(y_{1:T}; \Theta_1, \Theta_2) = \sum_{t=1}^{T} \ln \left( \sum_{\ell=1}^{4} f(\epsilon_{t,\ell}; \Theta_1) \times \pi(\ell|t-1; \Theta_2) \right)
$$

---

\(^8\) Equation (1.6) implies that $1 \cdot \text{Prob}[q_t = +1|q_{t-1}] - 1 \cdot \left[1 - \text{Prob}[q_t = +1|q_{t-1}]\right] = q(1 - \rho) + \rho q_{t-1}$. It then follows that $\text{Prob}[q_t = +1|q_{t-1}] = \frac{1}{2} \left(1 + q(1 - \rho) + \rho q_{t-1}\right)$.  

---
where $f(\epsilon_t; \Theta_1)$ is the normal density with mean 0 and variance $\Sigma_{\epsilon_t}$; with $\pi_\ell(t|t-1; \Theta_2)$ = \textbf{Prob} $[ A_t = \tilde{A}_\ell | \psi_{t-1}]$ being the probability of the measurement matrix taking particular values given by equation (1.15). Evaluating the likelihood requires computation of the predicted probabilities $\pi_\ell(t|t-1; \Theta_2)$ which in turn require computing the filtered probabilities $\pi_\ell(t|t; \Theta_2, \psi_t)$ after observing the data $y_t = \Delta p_t$. Collecting the predicted and filtered probabilities into $(4 \times 1)$ vectors $\xi_{t|t-1}$ and $\xi_{t|t}$, respectively, and the respective densities for each $\ell$ at time $t$ in the $(4 \times 1)$ vector $\eta_t = f(\epsilon_t; \Theta_1 | A_\ell, \psi_{t-1})$, the updates are accomplished as follows:  

\[
\begin{align*}
\hat{x}_{t|t-1} &= \hat{x}_{t-1|t-1} \hat{P} \\
\hat{x}_t &= \hat{x}_{t|t-1} \otimes \eta_t
\end{align*}
\]  

(1.19)  

(1.20)

where $\hat{P}$ defined by (1.10), $\otimes$ is the Hadamard product and the summation is over $4$ states. I now establish the recursions for the filters associated with the state $x_t$; with those for the switching process $A_t$ remaining as defined in (1.19).

The predictors $x_{t|t-1} = \mathbb{E} [x_t | y_{1:t-1}]$ and filters $x_t = \mathbb{E} [x_t | y_{1:t}]$, and their associated error variance–covariance matrices, respectively denoted by $P_{t|t-1} = \mathbb{E} \left\{ (x_t - x_{t|t-1}) (x_t - x_{t|t-1})' \right\}$ and $P_t = \text{Cov} \left\{ x_t | y_{1:t-1}, \epsilon_t \right\}$, are given by:

\[
\begin{align*}
x_{t|t-1} &= \Phi x_{t-1} \\
\hat{P}_{t|t-1} &= \Phi \hat{P}_{t-1} \Phi' + \Sigma_u \\
x_t &= x_{t|t-1} + \sum_{\ell=1}^4 \pi_\ell(t|t) K_{t,\ell} (y_t - \tilde{A}_\ell x_{t|t-1}) \\
P_{t|t} &= \sum_{\ell=1}^4 \pi_\ell(t|t) (I - K_{t,\ell} \tilde{A}_\ell) \hat{P}_{t|t-1} \\
K_{t,\ell} &= \hat{P}_{t|t-1} \tilde{A}_\ell \Sigma_{\epsilon_{t,\ell}}^{-1}
\end{align*}
\]  

(1.21)  

(1.22)  

(1.23)  

(1.24)  

(1.25)

where $\Sigma_{\epsilon_{t,\ell}} = \text{Var}(\epsilon_{t,\ell}) = \tilde{A}_\ell \hat{P}_{t|t-1} \tilde{A}_\ell'$ is the variance of the innovation $\epsilon_{t,\ell} = (y_t - \tilde{A}_\ell x_{t|t-1})$ and (1.25) is the Kalman gain. Equations (1.21) and (1.22) are standard in the filtering literature. Equations (1.23) and (1.24) exhibit the filter values as weighted linear combinations of the innovation values $\epsilon_{t,\ell}$ and $\Sigma_{\epsilon_{t,\ell}}$ corresponding to the four measurement matrices $\tilde{A}_\ell, \ell = 1,2,3,4$. These equations are similar to the approximations first introduced by Bar-Shalom and Tse (1975); Masreliez and Martin (1977) and Peña and Guttmann (1989). Following Shumway and Stoffer (2011, Sec 6.10), equation (1.23) can be confirmed by letting

---

the indicator $I(A_t = \bar{A}_\ell) = 1$ when $A_t = \bar{A}_\ell$ and zero otherwise. Then

\[
x_t = \mathbb{E}(x_t|y_{1:t}) = \mathbb{E}\left[\mathbb{E}(x_t|y_{1:t}, A_t) | y_{1:t}\right]
\]

\[
= \mathbb{E}\left\{\sum_{\ell=1}^4 \mathbb{E}(x_t|y_{1:t}, A_t = \bar{A}_\ell) I(A_t = \bar{A}_\ell) | y_{1:t}\right\}
\]

\[
= \mathbb{E}\left\{\sum_{\ell=1}^4 \left[x_t^{\ell-1} + K_{t,\ell} (y_t - \bar{A}_\ell x_t^{\ell-1})\right] I(A_t = \bar{A}_\ell) | y_{1:t}\right\}
\]

\[
= \sum_{\ell=1}^4 \pi_{t}(t|t) \left[x_t^{\ell-1} + K_{t,\ell} (y_t - \bar{A}_\ell x_t^{\ell-1})\right] = x_t^{\ell-1} + \sum_{\ell=1}^4 \pi_{t}(t|t) K_{t,\ell} (y_t - \bar{A}_\ell x_t^{\ell-1})
\]

Equation (1.24) may be confirmed similarly. This completes the specification of the data needed to evaluate the likelihood function and estimate parameters. Before embarking on an implementation of the model, I review alternative estimators commonly used in the literature. These alternates will later be used as a benchmark to compare my results to those of some easily implemented estimators.

### 1.3 Alternative Spread Estimators

The basic Roll model defined by (1.1) with a constant bid-ask spread implies the following dynamics for transaction prices:

\[
\Delta p_t = \Delta m_t + s \Delta q_t = s \Delta q_t + u_t^m
\]

from which it follows that the spread can be estimated as: $s = \sqrt{-\text{Cov}(\Delta p_t, \Delta p_{t-1})}$ where $\text{Cov}(\cdot, \cdot)$ is the auto–covariance of price changes. The estimate of $s$ obtained from the auto-covariance is usually called the moment estimator because it uses the sample estimate for the unknown population parameter. Several versions of this model have been used to infer the bid-ask spread and market liquidity.\(^{10}\) The moment estimator of equation (1.26) is feasible only if the sample auto covariance is negative, otherwise it is undefined (Harris, 1990). To overcome this difficulty, several methods have been developed to obtain estimates that do not rely on the sample auto–covariance.

In the next subsections, I review two methods that have been used to estimate the spread: the Gibbs sampling approach of Hasbrouck (2004, 2009) and two low frequency estimators: (i) the non-parametric method of Corwin and Schultz (2012) based on daily High–Low prices\(^{11}\) and (ii) the moment like approach of Abdi and Ranaldo (2017) based on daily Close–High–Low prices. Other methods and models have been used to estimate the bid-ask spread from daily data. These include estimators derived from transaction

---

\(^{10}\)The basic model is usually augmented with trading volume (in a regression framework) to additionally estimate the market impact of a trade (Hasbrouck, 2004; Vayanos and Wang, 2012).

\(^{11}\)See also Bleaney and Li (2016); Liu, Luo and Zhao (2016); Li, Lambe and Adegbite (2018) for variants of these non–parametric spread estimators with applications. These estimated proxies of the bid–ask spread have been used in studies based on daily data and with samples covering periods as far back as the 1930s (Pástor and Stambaugh, 2003).
price tick sizes (Holden, 2009; Goyenko, Holden and Trzcinka, 2009) and estimators derived from
the frequency of returns (Lesmond, Ogden and Trzcinka, 1999). Chen, Linton and Yi (2017)
develop a method for estimating the spread based on the empirical characteristic function instead of Roll’s autocovariance.

1.3. Alternative Spread Estimators

1.3.1 The Gibbs Sampling Estimator

The Gibbs sampling approach of Hasbrouck (2002) has been applied in many research papers (see Hasbrouck, 2009; Schestag, Schuster and Uhrig-Homburg, 2016; Liu, Luo and Zhao, 2016, amongst others). It involves using daily trade by trade transaction prices and simulating the trade direction series and the spread defined in (1.1). The price change defined in (1.26) imply the variance and auto-covariance of price changes are given by:

\[ \gamma_0 \equiv \text{Var}(\Delta p_t) = \sigma_m^2 + s^2 \text{Var}(\Delta q_t) = \sigma_m^2 + 2s^2 \]

\[ \gamma_1 \equiv \text{Cov}(\Delta p_t, \Delta p_{t-1}) = s^2 \mathbb{E}(\Delta q_t, \Delta q_{t-1}) = -s^2 \]

Solving (1.27) gives \( \sigma_m^2 = \gamma_0 + 2\gamma_1 \) and \( s = \sqrt{-\gamma_1} \). The Gibbs sampler involves two steps:

i. Parameter draw: generate random draw of \( s, \sigma_m \mid p, m, q \).

ii. Latent data draw: generate a random draw of \( m, q \mid s, \sigma_m, p \).

The algorithm involves conditioning on the most recent values of the latent data when making parameter draws and conditioning on the most recent parameter values when making latent data draws. The classical Bayesian regression with an explained variable \( y \) and regressors \( X \) is \( y = X\beta + u \). If the prior distribution on the coefficients is normal: \( \beta \sim N(\mu^{Prior}_\beta, \Omega^{Prior}_\beta) \) then the posterior distribution of coefficients is \( \beta \mid y \sim N(\mu^{Pstr}_\beta, \Omega^{Pstr}_\beta) \), where \( \mu^{Pstr}_\beta = Dd \) and \( \Omega^{Pstr}_\beta = D^{-1} = X'\Omega_u^{-1}X + \Omega^{Prior}_\beta^{-1} \) with \( \mu^{Prior}_\beta, \Omega^{Prior}_\beta \).

Mapping the standard Bayesian regression into the Roll model, \( y = [\Delta p_t], X = [\Delta q_t] \), \( \beta = s \) and \( \Omega = \sigma_m^2 \). Analogous to the standard Bayesian regression, the prior for \( s \) can be set to normal with \( \mu^{Prior}_s = 0 \) and \( \Omega^{Prior}_s = 10^6 \) as in Hasbrouck (2004). A non-negativity constraint may be imposed on \( s \) so that the prior is \( s \sim N(0, \Omega^{Prior}_s)^+ \). The + superscript denotes restriction of the density to a domain \([0, +\infty)\) with the posterior \( N(\mu^{Pstr}_s, \Omega^{Pstr}_s) \) from which \( s \) is drawn. Given a normal \( u^m \), the prior for \( \sigma_m^2 \) is the inverted gamma distribution.

1.3.2 Low Frequency Estimators

The Corwin-Schutz High–Low (HL) Estimator

Corwin and Schultz (2012) propose an estimator of the bid-ask spread based on the daily

\[ \text{The positive domain half-normal density is defined by: } \text{ } X \sim N(0, \sigma^2), \text{ } Y = |X| \text{ and } f_Y(y) = \frac{\sigma}{\sqrt{2\pi}} \exp \left( -\frac{y^2}{2\sigma^2} \right), y > 0. \]

\[ \text{In implementation, the response } y \text{ is parameterized in terms of the mean and precision, } y \sim N(0, |\text{Var}(y)|^{-1}), \text{ so an inverse gamma prior for the variance is actually a gamma prior.} \]
high and low transaction prices over any two days period. Their spread estimator, denoted $s_{HL}$, is computed as:

$$s_{HL} = \frac{2(e^\alpha - 1)}{1 + e^\alpha}, \quad \text{where} \quad \alpha = \sqrt{\frac{2\beta - \sqrt{\beta}}{3 - 2\sqrt{2}}} - \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}} \quad (1.28)$$

with $\beta$ and $\gamma$ in the expression for $\alpha$ in equation (1.28) given by:

$$\beta = \mathbb{E}\left[\sum_{j=0}^{1} \ln \left(\frac{H_{t+j}^0}{L_{t+j}^0}\right)\right], \quad \gamma = \mathbb{E}\left[\ln \left(\frac{H_{t+1}^0}{L_{t+1}^0}\right)\right]^2.$$

$H_{t}^0$ and $L_{t}^0$ are the observed high(low) stock price of day $t$ while $H_{t,t+1}^0$ and $L_{t,t+1}^0$ are the observed high(low) prices over days $t$ and $t+1$.

The Abdi-Ranaldo Close–High–Low (CHL) Estimator

Abdi and Ranaldo (2017) have developed a low frequency estimator of the spread based on readily available daily close, high and low prices (CHL). Their estimator is similar to a moment based estimator of the basic Roll model. To derive their estimator, they begin by assuming that the transactions price process follows the basic Roll model of equation (1.1) with the daily high price ($h_t$) being buyer initiated ($q_t^h = 1$) while the daily low price ($l_t$) is seller initiated ($q_t^l = -1$):

$$c_t = m_t^e + q_t \frac{s}{2}, \quad q_t = \pm 1, \quad h_t = h_t^e + \frac{s}{2}, \quad l_t = l_t^e - \frac{s}{2}$$

where $c_t$ is the daily close price and $m_t^e$ is the value of the efficient price at closing time. They further define the mid-range as the average of daily high and low log-prices, which coincides with the mid-range of the efficient price:

$$\eta_t = \frac{l_t + h_t}{2}.$$

The efficient price hits $\eta_t$ at least once a day. They argue that an unbiased proxy for the end-of-the-day midquote for day $t$, is the average of the mid-ranges of over dates $t$ and $t+1$, because the end-of-the-day midquote of day $t$ falls within the time window over which $\eta_t$ and $\eta_{t+1}$ are hit. The squared difference between the close price on day $t$ and the midpoint proxy include a bid-price component and the variance of the efficient price process:

$$\mathbb{E}\left[(c_t - \frac{\eta_t + \eta_{t+1}}{2})^2\right] = \frac{s^2}{4} + \frac{1}{2} \left(1 - \ln(2)\right)\sigma_u^2$$

Since mid-ranges are independent of the spread, their difference reflects the volatility of the efficient price process:

$$\mathbb{E}\left[(\eta_{t+1} - \eta_t)^2\right] = 2\left(1 - \ln(2)\right)\sigma_u^2$$

17
1.4 Simulation Study

The expressions for the squared difference between prices and the difference between mid-ranges then gives the following expression for the spread:

\[ s_{\text{CHL}}^2 = 4 \mathbb{E} \left[ (c_t - \eta_t) (c_{t+1} - \eta_{t+1}) \right]. \]  

(1.29)

1.4 Simulation Study

In this section, I give results of a simulation study that uses artificial data to gauge how well the model performs in comparison to standard approaches found in the literature. For the simulation studies, I first present results of my model implementation in comparison to the Gibbs sampling estimates. This initial stage of evaluating the model assumes that all trades are observed as they sequentially occur throughout the day for 100 days. This is followed by a low frequency evaluation over 62 (3 months), 126 (6 months) and 253 (1 year) days of trading but under the assumption that the data available are the daily open, high, low and close prices. This mirrors the data commonly used in the implementation of low-frequency estimators such as by Abdi and Ranaldo (2017) and Corwin and Schultz (2012).

1.4.1 High Frequency

I perform simulation studies by generating artificial data following the approach of Hasbrouck (2004) and Chen et al. (2017). The simulation studies assume that log-prices are generated by the basic Roll model in equation (1.1), with \( m_0 = 100, \sigma_{m}^2 = 0.01^2 \) and \( s_t = s = 0.01 \) in each day, for 100 days of trading. The number of trades in a day is drawn from a discrete normal distribution over the set \{15, 16, ..., 25\}. This gives 1,981 observations with a median of 20 trades per day. Figure 1.1 shows the generated times series for the 100 days of trading (upper panel) and the simulated trade direction indicator \( q_t \) together with prices (lower panel) for the first 82 trades over 3 days of trading with the number of trades 23, 15 and 21. To obtain the Gibbs sampling estimators, I use the MATLAB code accompanying Hasbrouck (2004) from the author’s website (retrieved on Oct 17, 2016) and use 10,000 sweeps of the Gibbs sampler with a burn-in of 2,000.

Table 1.1 gives parameter estimates using the full sample using the switching regression equation (1.13) in column (1) and (1.18) in columns (2)–(3). The model replicates the simulated data parameters as expected: \( s = 0.0096 \approx 0.01 \) and \( \sigma_{m} = 0.01 \). The estimated transition probabilities in column (1) imply that the unconditional probability of a buyer or seller initiated trade are approximately equal: \( \pi_1 = \text{Prob}(q_t = +1) = \frac{p_{21}}{p_{12}+p_{21}} \approx 0.5 \approx \pi_2 \). \( \sigma_{q} \) is statistically not different from zero as we would expect given that the generated data use a constant \( s_t = s \). In columns (2) and (3), I have dispensed with the switching structure and instead used the linear state space model defined by equation (1.18). The parameters \( s, \sigma_{m}, \sigma_{q} \) are all precisely estimated: \( \sigma_{q} = 4 \pi_1 \pi_2 \) should be approximately equal to one and \( \rho = 1 - p_{12} - p_{21} \) should be zero. The estimates for \( q \) and \( \sigma_{q} \) exhibit large standard
errors, – because $q = \mathbb{E}(q_t) = \pi_1 - \pi_2 \approx 0$ and should be dropped from the estimation. This is done in column (3) with $q$ and $\sigma_s$ set to zero.

**Comparison to Gibbs Sampling Estimator** Table 1.2 gives a summary of the time series of spreads generated by the model estimates in Table 1.1 and those using the Gibbs sampling estimator. The rows labeled with a capital $S$ as the level version the spread: $S = \bar{P} \times (e^{s_t} - e^{-s_t})$ where $\bar{P}$ (in ticks) is the average transaction price for the length of the simulation. The summaries show that the parametric model used here is very precise, with results similar to those obtained by Hasbrouck’s (2004) Gibbs sampling estimator. Figure 1.2 shows the distribution of these time series. The estimated $s_t$ are visually well defined, unimodal and concentrated.

**Classification Performance** In the financial markets microstructure literature, the tick rule (TR), Lee-Ready (LR) algorithm and the bulk volume classification (BVC) are frequently used methods for labeling trades as either buyer or seller initiated (Chakrabarty, Pascual and Shkilko, 2015; Easley, de Prado and O’Hara, 2016). The tick rule assigns a trade to be a buy if the trade price was an uptick relative to the previous trade and to be a sell if it was a downtick. In case of a zero-tick, the movement relative to the last price change is used. The LR algorithm uses a quote rule and the tick rule to sign trades. The quote rule classifies a trade as buyer (seller) initiated if the trade above (below) the midpoint. The TR is then used to sign trades that occur at the midpoint (Lee and Ready, 1991). The BVC aggregates trading activity over time, volume, or trade intervals (bars) and then uses the standardized price change between the bars to assign a fraction of the volume as buyer-
1.4. Simulation Study

Table 1.1: Parameter Estimates for High Frequency Data

<table>
<thead>
<tr>
<th>Model</th>
<th>(1) $q_t = \pm 1$</th>
<th>(2) $q_t = \text{AR}(1)$</th>
<th>(3) $q_t = \text{AR}(1) \mid q = 0, \sigma_s = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Estimate</td>
<td>S.E.</td>
<td>Estimate</td>
</tr>
<tr>
<td>$s$</td>
<td>0.0096</td>
<td>0.0003</td>
<td>0.0098</td>
</tr>
<tr>
<td>$\sigma_m$</td>
<td>0.0104</td>
<td>0.0005</td>
<td>0.0103</td>
</tr>
<tr>
<td>$\sigma_s$</td>
<td>0.0013</td>
<td>0.0022</td>
<td>0.0027</td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>0.5277</td>
<td>0.0330</td>
<td>–</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.4896</td>
<td>0.0340</td>
<td>–</td>
</tr>
<tr>
<td>$\sigma_q$</td>
<td>–</td>
<td>–</td>
<td>0.9917</td>
</tr>
<tr>
<td>$q$</td>
<td>–</td>
<td>–</td>
<td>-0.2678</td>
</tr>
<tr>
<td>$\rho$</td>
<td>–</td>
<td>–</td>
<td>0.0230</td>
</tr>
<tr>
<td>LogL</td>
<td>-5117.5629</td>
<td>–</td>
<td>-5096.3872</td>
</tr>
<tr>
<td>$T$</td>
<td>1881</td>
<td>–</td>
<td>1881</td>
</tr>
<tr>
<td>AIC</td>
<td>5.4466</td>
<td>5.4252</td>
<td>5.4231</td>
</tr>
</tbody>
</table>

Table 1.2: Distribution of High Frequency Spread Estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Med.</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spreads $s</td>
<td>q = \pm 1$</td>
<td>0.0096</td>
<td>0.0095</td>
<td>0.0001</td>
<td>0.0094</td>
</tr>
<tr>
<td>$s</td>
<td>q = \text{AR}(1)$</td>
<td>0.0098</td>
<td>0.0098</td>
<td>0.0001</td>
<td>0.0094</td>
</tr>
<tr>
<td>$s$ Gibbs</td>
<td>0.0097</td>
<td>0.0097</td>
<td>0.0002</td>
<td>0.0089</td>
<td>0.0104</td>
</tr>
<tr>
<td>Ticks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S</td>
<td>q = \pm 1$</td>
<td>1.9161</td>
<td>1.9101</td>
<td>0.0203</td>
<td>1.8878</td>
</tr>
<tr>
<td>$S</td>
<td>q = \text{AR}(1)$</td>
<td>1.9647</td>
<td>1.9642</td>
<td>0.0239</td>
<td>1.8886</td>
</tr>
<tr>
<td>$S$ Gibbs</td>
<td>1.9409</td>
<td>1.9426</td>
<td>0.0477</td>
<td>1.7743</td>
<td>2.0781</td>
</tr>
</tbody>
</table>

initiated (Chakrabarty et al., 2015). To an approximation, the low-frequency version of the model used here is similar to the BVC, which is probabilistic (see the discussion by Easley et al., 2016, Section 2.2).

One outcome of the switching model is the probabilistic inference on the trade direction/initiator indicator. For the whole sample of 1881 trades, the filtered probabilities of equation (1.19) correctly label 76% of trades using the rule $q_t = +1$ if \(\text{Prob}(q_t = +1 | \psi_{t-1}) > \frac{1}{2}\). To get better performance in labeling trades, I drop the first day observations and reduce the sample to 5 days of trading, with 91 observations. This increases the correctly labeled trades to 82.42%. For the same observations, an application of the tick rule predicts 65.93% of the observations correctly. This is prediction accuracy for the TR is around the value found by Blais and Protter (2012) for a highly liquid security. Figure 1.3 shows the probabilities for the first 40 observations used in the exercise, where just 2 trade direction indicators are about indeterminate. This performance is comparable to those of the LR, TR and BVC models (Ellis, Michaely and O’Hara, 2000; Odders-White, 2000) or better in comparison to methods involving some form of estimation (see for example Blazejew-
ski and Coggins, 2005, who achieve an average out-of-sample classification accuracy of 71.40% using a k-nearest neighbor with three predictor variables). I now move to present and discuss the more interesting results from low-frequency estimation.

**Figure 1.3:** Labeling Accuracy: High Frequency

---

**1.4.2 Low Frequency**

I estimate the model specified by equation (1.16). To perform the estimation, I simulate 254 days of trading (one year), assuming that the first day log-prices are generated by the basic Roll model in equation (1.1), with \( m_{1,1} = 100, \sigma_m^2 = 0.01^2 \) and \( s_{1,1} = s = 0.01 \). In each subsequent day, I set \( m_t,1 = m_{t-1,n} + u'^n \) and \( s_t,j = s = 0.01 \) for each day \( t = 2, \ldots, 254 \) and every trade \( j = 1, \ldots, n_t \), where \( n_t \) is the number of trades in a day, drawn from a discrete normal distribution over the set \{15, 16, \ldots, 25\}. This gives 254 observations with a median of 20 trades per day. On each simulated day, I collect the daily open, close, high, low.
1.4. Simulation Study

collect the efficient prices at open and close, \( m_{t,1} \) and \( m_{t,n} \) as well as the trade direction indicators \( q_{t,1} \) and \( q_{t,n} \) for later comparison with the estimated values (these data are however not used in the estimation). Figure 1.4 shows the first 60 points of the simulated data.

**Figure 1.4: Daily Simulated Price Series**

Table 1.3 gives parameter estimates and standard errors for three low frequency models with different sample sizes: 62 (approximately one quarter of trading), 126 (6 months) and 252 (a year). The estimated model replicates the simulation parameters \( s \) and \( \sigma_m \) with \( \sigma_s \approx 0 \) in all cases. The parameters \( p_{11}^{(n)} \) and \( p_{22}^{(n)} \) define the transition matrix \( \mathbf{P}^{(n)} \). These are estimated instead of using the definition of \( n \)-step transition given by equation (1.4). The estimation converges faster when they are estimated rather than redefined using equation (1.4).

**Table 1.3: Parameter Estimates for Low Frequency Data**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( T = 62 )</th>
<th>( T = 126 )</th>
<th>( T = 253 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s )</td>
<td>0.0085 0.0023</td>
<td>0.0095 0.0013</td>
<td>0.0092 0.0012</td>
</tr>
<tr>
<td>( \sigma_m )</td>
<td>0.0091 0.0023</td>
<td>0.0094 0.0014</td>
<td>0.0096 0.0013</td>
</tr>
<tr>
<td>( \sigma_s )</td>
<td>0.0000 0.0000</td>
<td>0.0000 0.0000</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>( p_{11} )</td>
<td>0.4647 0.2495</td>
<td>0.4119 0.2871</td>
<td>0.5446 0.9679</td>
</tr>
<tr>
<td>( p_{22} )</td>
<td>0.0000 0.0073</td>
<td>0.5187 0.2865</td>
<td>0.1980 5.6362</td>
</tr>
<tr>
<td>( p_{11}^{(n)} )</td>
<td>0.7239 0.1733</td>
<td>0.6465 0.2261</td>
<td>0.6931 1.4270</td>
</tr>
<tr>
<td>( p_{22}^{(n)} )</td>
<td>0.1456 0.1194</td>
<td>0.5456 0.2650</td>
<td>0.3459 1.6377</td>
</tr>
<tr>
<td>LogL</td>
<td>−169.5470</td>
<td>−338.8867</td>
<td>−677.6234</td>
</tr>
<tr>
<td>Label Accuracy</td>
<td>74%</td>
<td>75%</td>
<td>72%</td>
</tr>
</tbody>
</table>

Figure 1.5 shows the probabilities associated with \( q_t = +1, s_t = s \) for the model in column (2) of Table 1.3. As is clearly visible, the model does not allocate as high probabilities to trade labels as it does in the high frequency version shown in Figure 1.3. This is to be expected given the loss of data points in the low frequency model. However, using the rule \( q_t = +1 \) if \( \text{Prob}(q_t = +1|\psi_{t-1}) > \frac{1}{2} \) as before, the labeling success does not reduce by much as shown
in last row of Table 1.3. However, the larger the sample used, the more challenging it is to obtain accurate labeling. Figure 1.6 shows the estimated underlying price process at the open $\hat{m}_{t,1}$ to its true value for the model in column (2) Table 1.3. The estimate closely tracks the true value, in line with the closeness of the close proxy $\hat{m}_{t,n1}$ to the value of underlying at close $m_{t,n}$.

**Figure 1.5: Labeling Accuracy: Low Frequency**

![Figure 1.5: Labeling Accuracy: Low Frequency](image)

**Figure 1.6: Estimated Underlying Price Process at Open: $\hat{m}_{t,1}$**

![Figure 1.6: Estimated Underlying Price Process at Open: $\hat{m}_{t,1}$](image)

1.5 Empirical Applications

1.5.1 Data Description

The prices I study consist of three commodity futures contracts from three sectors: Coffee (Agriculture), Natural Gas (Energy) and Palladium (Metals). The data consists of continuous contracts for each commodity starting from 11 October 2015 (longer time series are available; the choice of the start date is arbitrary). The data come from two exchanges: Intercontinental Exchange (ICE) and the Chicago Mercantile Exchange (CME). I use the daily settlement price (close), open, high and low prices. The procedure for determining the settlement price may vary across commodities, but it generally involves determining the weighted average price during a “closing range” period (Marshall, Nguyen and Visaltanachoti, 2012). Table 1.4 gives a summary of the descriptive statistics and Figure 1.7 depicts the first 3 months of daily prices of the commodity futures considered. These show the similarity in movement with the simulated low frequency prices shown in Figure 1.4.
1.5. Empirical Applications

**Figure 1.7:** Open, Close, Min and Max Futures Prices (3 Months)

(a) Coffee
(b) Natural Gas
(c) Palladium

1.5.2 Estimation Results

Table 2.2 gives parameter estimates for the three commodities. The estimate $s = 0.0006$ for Natural Gas implies that the spread is equal to $\hat{P} \times (e^{0.0006} - e^{-0.0006}) = 2.33 \times 0.0012$. The Natural Gas NYMEX futures contract trades in units of million British thermal units (mmBtu). The contract size is 10,000 mmBtu with a minimum Tick Size: $0.001$ per mmBtu, worth $10.00$ per contract. The estimated log-half spread of 0.0006 is very close to the minimum tick of the contract: $e^{0.0006} - e^{-0.0006} = 0.0012$. Given the average price of 2.33, the bid–ask spread is $0.0012 \times 2.33 \times 10,000 = 27.96$ or about 3 ticks per contract. The same applies to the log-half spread estimate for Palladium. The metal trades in contract units of 100 troy ounces with price quotations in U.S. dollars and cents per troy ounce. The minimum price fluctuation 0.10 per troy ounce $\equiv$ $10.00$. The estimate $s = 0.016$ in col-
Table 1.4: Descriptive Statistics of Sample Prices

<table>
<thead>
<tr>
<th>Contract</th>
<th>Symbol</th>
<th>T</th>
<th>Start</th>
<th>Prices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean Min Max Std. Dev.</td>
</tr>
<tr>
<td>Coffee</td>
<td>KC</td>
<td>252</td>
<td>11/10/15</td>
<td>Close 129.20 111.60 156.55 12.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Open 129.21 112.00 156.60 12.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High 130.54 113.35 160.90 12.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low 127.71 111.05 156.00 12.09</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>NG</td>
<td>252</td>
<td>11/10/15</td>
<td>Close 2.33 1.64 3.27 0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Open 2.33 1.63 3.19 0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High 2.38 1.68 3.29 0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low 2.29 1.61 3.15 0.38</td>
</tr>
<tr>
<td>Palladium</td>
<td>PA</td>
<td>252</td>
<td>11/10/15</td>
<td>Close 595.58 469.80 726.40 70.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Open 594.87 468.20 725.85 72.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High 600.59 482.00 747.50 72.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low 588.97 451.50 714.55 71.14</td>
</tr>
</tbody>
</table>

umn (3) of Table 1.5 and the average close price $\bar{P} = 595$ from Table 1.4 then means that the spread equals $595 \times (e^{0.016} - e^{-0.016}) = 19.0408$ or approximately 2 ticks. The ICE coffee futures contract, with symbols KT at CME Globex or KC at the ICE, is quoted in U.S. dollars and cents per pound, with a contract for 37,500 pounds and a minimum tick size of $0.0005$ per pound ($18.75$ per contract $= 37,500 \times 0.0005$). The estimate of $\sigma = 0.001$ for coffee then implies $e^{0.001} - e^{-0.001} = 0.002$ so the spread equals $0.002/0.0005 = 4$ ticks, which is to be expected for a small tick asset.

The transition probability estimates $p_{11}, p_{22}$ of the Coffee and Natural Gas contracts imply that the unconditional probability of a buyer initiated trade are approximately equal: $p_1 = \text{Prob}(q_{t,1} = +1) = \frac{p_{11}}{p_{11} + p_{21}} = \frac{1-0.9474}{(1-0.9436) + (1-0.0474)} = 0.99, p_2 = 0$, so all trades at the open are buyer initiated. A similar computation for probabilities at close shows that almost all transactions are invariably seller initiated. This is consistent with market markers (or dealers) initiating trades to control excess inventory (Locke and Sarajoti, 2004) and day traders or short sellers closing their positions at the end of trading. For the palladium contract, we can compute $p_1 = \frac{(1-0.9436)}{(1-0.9436) + (1-0.9750)} = 0.69, p_2 = 0.31$ at open and approximate the order flow process of equation 1.7: $q_{t,1} = q(1-\rho) + \rho q_{t-1,n} + u_{(t,1)}^Q$, using $q = p_1 - p_2 = 0.38$ and $\rho = 1 - ((1-0.9436) + (1-0.975)) = 0.9186$. Forecasting the first order of the day requires knowledge of $q_{t-1,n}$ which we can approximate using the parameters at close: $p_{11}^{(n)}$ and $p_{22}^{(n)}$.  

25
### 1.5. Empirical Applications

**Table 1.5: Parameter Estimates: Commodity Prices**

<table>
<thead>
<tr>
<th>Commodity</th>
<th>Parameter</th>
<th>Estimate (1)</th>
<th>S.E. (1)</th>
<th>Estimate (2)</th>
<th>S.E. (2)</th>
<th>Estimate (3)</th>
<th>S.E. (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee</td>
<td>$s$</td>
<td>0.0010</td>
<td>0.0004</td>
<td>0.0006</td>
<td>0.0005</td>
<td>0.0160</td>
<td>0.0019</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>$\sigma_m$</td>
<td>0.0048</td>
<td>0.0010</td>
<td>0.0161</td>
<td>0.0007</td>
<td>0.0012</td>
<td>0.0005</td>
</tr>
<tr>
<td>Palladium</td>
<td>$\sigma_s$</td>
<td>0.0046</td>
<td>0.0008</td>
<td>0.0000</td>
<td>16.9792</td>
<td>0.0077</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>$p_{11}$</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.9750</td>
<td>0.0151</td>
</tr>
<tr>
<td></td>
<td>$p_{22}$</td>
<td>0.0474</td>
<td>0.1020</td>
<td>0.0000</td>
<td>0.0253</td>
<td>0.9436</td>
<td>0.0244</td>
</tr>
<tr>
<td></td>
<td>$p_{11}^{(n)}$</td>
<td>0.2649</td>
<td>0.1189</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1731</td>
<td>0.0379</td>
</tr>
<tr>
<td></td>
<td>$p_{22}^{(n)}$</td>
<td>1.0000</td>
<td>0.0101</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

*Model: $q_t = \pm 1, s_t = s + u_t^s, \ u_t^s \sim N(0, \sigma_s^2)$*

**1.5.3 Comparison to Daily High–Low (HL)**

Figures 1.8a–1.8c show time series plots of the filtered log–half spreads against the daily High–Low (HL) and Close–High–Low (CHL) estimators of Corwin and Schultz (2012) and Abdi and Ranaldo (2017) described in Sec. 1.3.2. The HL/CHL estimates show remarkable daily variation that are generally difficult to reconcile with stable day–to–day market conditions. In many days, the HL estimate also gives negative values of spread estimates, which should not be observed under any normal market conditions. The model proposed here gives smooth yet time–varying spread estimates that are rarely negative.

---

14Corwin and Schultz propose various methods of dealing with the negative estimates.
1.6 Conclusion

This paper has developed a new parametric method for the estimation of transaction costs using only observed trade prices without requiring knowledge of the initiator side of the trades. In classification of observed prices as buyer or seller initiated, the method performs better than the popular tick rule. In estimation of the effective bid–ask spread, the method works well for both high–frequency trade–by–trade data and for low–frequency daily summaries of trading activity while also being applicable with small samples.

From a practical perspective, the model specification allows for real time approximation and forecasting of transaction costs, without the dedicated computing infrastructure that large data sets such as limit order books would require. The estimates of transaction costs may be used by an agent trading in the financial markets to minimize the execution
1.6. Conclusion

costs of a liquidation or portfolio balancing activity. Finally, the probabilistic inference obtained as a natural byproduct of the estimation may be used to approximate the best trading times as they give a preview of the order flow process.
CHAPTER 2

Price impact as reaction to order flow imbalance

Abstract

We postulate a theory of transaction history dependent price formation in a limit order book market for commodity futures. We hypothesize that trading agents post mid-price updates equal to half the bid–ask spread in response to the observed sequence of buy and sell orders. The theory leads to falsifiable empirical predictions of the relationship between price impact of a trade, the effective spread and volatility. In empirical tests using tick data from the Tokyo Commodities Exchange, we find that a large fraction of price volatility can be explained by the correlation of order signs through a given trading session and conclude that high frequency price dynamics are driven by reactions to the order flow imbalance.

2.1 Introduction

Most modern financial markets operate limit order books (LOBs), where financial institutions interact via the submission of orders. A buy order (sell order) is a commitment to buy (sell, respectively) a maximum quantity of an asset for a price no larger than its limit price (no smaller than its limit price, respectively). There are two frameworks of thinking about how agents arrive at the prices of the orders they place on the LOB. The predominant framework in economics is the efficient markets hypothesis according to which “agents successfully forecast short-term price movements and trade accordingly” (Bouchaud, Bonart, Donier and Gould, 2018, pg. 209). A trader who anticipates a rise in prices is likely to make a buy order, resulting into a positive correlation between the sign of trades and the subsequent price moves. Trade reveals information about the fundamental value through the “price discovery process”.

An alternative framework is that of price impact: on average, the arrival of a buy trade causes prices to rise and the arrival of a sell trade causes prices to fall such that price updates are a reaction to the order flow imbalance. This theory asserts that at least in short time scales, a transaction will have an impact on prices even when it is not based on any information about fundamental value. While both frameworks result into positive correlation between trades signs and price movements, they are conceptually different: in the first, value drives trading while in the second, the act of trading itself impacts prices.
2.1. Introduction

For physically delivered commodity futures contracts, the dichotomy between the price discovery and impact frameworks may not be that clear cut since the contracts may reflect changes to future demand or supply of the underlying. For instance, the sale of a newly created contract by the producer of a commodity clearly transmits information about the possible future supply of the commodity – hence the trade in and of itself is the source of fundamental information. The same applies to the case where the end user of the commodity makes a transaction. We propose a history dependent price update system for LOB trading that takes these possibilities into account: price updates in response to observed order flow throughout a trading session. We assume that the transaction price consists of a mid-price plus a spread component, with seller(buyer) initiated trades occurring at the mid plus(resp. minus) half the bid-ask spread. We further assume that the sequence trade direction indicators (of buyer or seller initiated trades) are the outcome of a first order two-state hidden Markov process. Mid price changes are then driven by agents forecasts of the sign of the next transaction given the history of signs up to the current time. This simple framework leads to martingale prices with linear relationships between lag-1 and lag-\(k\) price impact functions, and between volatility per period and the average spread.

We use high frequency tick data from the Tokyo Commodities Exchange (TOCOM) to test the theory. We estimate a state–space system using transaction prices for two of the most liquid futures contracts traded at TOCOM on the 24th of April 2019: Gold and Platinum.\(^\text{15}\) We use the model to obtain estimates of the mid-price before a trade, bid-ask spreads and sign each transaction as either buyer or seller initiated. We then compute response functions to obtain an almost perfect relationship between lag-1 and lag-\(k\) impact and approximate the component of price volatility attributable to transaction costs and sign correlation. Our results strongly suggest that agents in financial markets make price updates in reaction to the observed order flow imbalance. Unlike the “sign surprise” models of Madhavan, Richardson and Roomans (1997), the agents in our market are proactive: price updates are made based on the prediction of the next transaction sign. This leads to a predictable variation in short term price changes directly as a result of the order flow, as long as the probability of transaction sign reversals is not too low.

The remainder of the article is organized as follows. In the next Section (2.2), we give detailed descriptions of the history dependent price formation process. Assumptions about the statistical properties of the spread, transactions sign generation process and the price update rule used by agents are given in Sec. 2.2.1. We then show in Sec. 2.2.2 that under these assumptions, the price process is a martingale and excludes quasi-arbitrage. Sub-section 2.2.3 revisits the concept of unconditional impact (response function) then derives two falsifiable empirical predictions about the linear relation between lag-1 and lag-\(k\) impact as well as an approximate formula that relates price volatility to the bid–ask

\(^{15}\)The choice of trading day, month and session is to ensure we use a relatively large sample of buy/sell transactions. Market activity varies by time of day, week or year and delivery month. The choice of 24th of April 2019 offers as a large enough sample of observations to make the analysis statistically valid.
spread and order flow. We then discuss these results in view of the literature in Sec. 2.2.4. Section 2.3 discusses our calibration strategy which involves the estimation of a non-linear state-space system. This is important for two reasons. First, the data we use do not contain any information about the sign of the transaction and/or the quoted spread which therefore need to be estimated. Second, the probabilistic nature of our model means that we need to make inference about the transition probabilities from buys to sells and vice-versa. The estimation allows us to obtain these probabilities and approximations of the mid-prices before a transaction which we use to perform tests of the empirical predictions in Section 2.4. Section 2.5 concludes.

2.2 Transactions history dependent prices formation

Roll (1984) suggested a simple model of security prices in the presence of transaction costs. Let \( p_t \) denote the transaction price at date \( t \) of a given security. Let \( m_t \) denote the efficient or mid price that would prevail in a frictionless market without any transaction costs. The dynamics of prices are:

\[
p_t = m_t + s_t q_t, \quad m_t = m_{t-1} + u_t^m, \quad s_t = s + u_t^s
\]

where \( s_t \) is the half bid-ask spread, \( q_t \) is a trade direction indicator:

\[
q_t = \begin{cases} 
+1 & \text{if transaction is buyer initiated} \\
-1 & \text{if transaction is seller initiated} 
\end{cases}
\]

and \( u_t^m \) is a zero-mean disturbance uncorrelated with \( q_t \). Roll motivates \( s_t \) as one-half the quoted bid-ask spread, but since the model refers to transaction prices, \( s_t \) can be viewed as the effective transaction cost (Hasbrouck, 2009). In equation (2.1), a buyer initiated transaction generates a transaction price that exceeds the efficient price by \( s_t \).

Equation (2.1) can be used to represent the best bid and ask prices in a LOB. When a buy or sell order is submitted, the LOB’s trade-matching algorithms check whether it can be matched against previously submitted but still unmatched sell (buy) orders, in which case a transaction occurs. If the order cannot be immediately matched, it remains active in the book until it is matched against a future incoming sell order or canceled. For a given LOB, the bid price is the highest price among active buy orders at time. Similarly, the ask price is the lowest price among active sell orders at time. The bid and ask prices are collectively known as the best quotes. Their difference is called the bid–ask spread, and their mean is the mid price. Figure 2.1 gives a schematic representation of a typical LOB.

Roll’s original specification was made under the following assumptions (Foucault, Pagano and Roell, 2013): (i) balanced order flow: market buy and sell orders are approximately of equal probability; (ii) no autocorrelation of orders – \( \text{Prob}(q_t = 1) = \text{Prob}(q_t = -1) = \frac{1}{2} \); (iii)
2.2. Transactions history dependent prices formation

Figure 2.1: Schematic Representation of a Limit Order Book

Notes: The horizontal lines within the blocks at each price level denote the different active orders at each price. Source: Bonart and Lillo (2018).

no effect of orders on the mid-quote – changes in the value of the security ($u_t^m$) and the direction of market orders ($q_t$) are independent; and finally (iv) non-varying fundamental expected return – the expected return is constant and equal to zero: $\mathbb{E}(m_t - m_{t-1}) = 0$.

Assumptions (i) and (ii) are unlikely to hold in actual markets as market orders do not occur with equal probability. For instance, buyer initiated trades are more likely at the end of a trading session when agents close their short positions. Whenever the transactions do not occur with equal probability, orders will show non-zero serial correlation. Assumptions (iii) and (iv) are problematic if some traders submitting market orders have information on the security’s future payoff or as we will later argue, trades have an effect on the future expected fundamental value. In the next Sec. 2.2.1, we relax these assumptions by specifying a first order Markov process for the order flow, allowing for serial correlation and transaction history to have an effect on the efficient price process – that is – permanent impact. The new assumptions generate a set of properties and new empirical predictions about the behavior of prices which we discuss in sections 2.2.2 and 2.2.3, respectively.

2.2.1 History dependent pricing

In order to model how transactions history affects the prices posted on the LOB, we make several assumptions regarding the behavior of the spread, trade signs and mid-price updates. Assumptions 1 and 2 simply capture the dynamics of prices without making any claims about the forces driving price changes. Assumption 3 is our theory about how
agents make price updates over a trading session. Together, the three assumptions lead to falsifiable predictions about the relationships between prices observed over any discrete interval of length $k \in \{1, 2, \ldots \}$.

**Assumption 1:** (Random Spread). Observed transaction prices $\{p_t\}_{t=1}^T$ are generated from equation (2.1) where $\{u^m_t, q_t\}_{t=1}^\infty$ is a strictly stationary process with $u^m_t \sim N(0, \sigma^2_m)$. The spread $s_t$ is a random variable that follows the process:

\[
s_t = s + u^r_t, \quad \text{where } u^r_t \sim N\left(0, \sigma^2_s\right) \quad \text{with} \quad u^r_t \perp u^m_t.
\]  

Equation (2.2) allows the spread to vary around a constant mean $\mathbb{E}[s_t] = s$ in response to changes in market conditions unrelated to the efficient price process (see e.g. Figure 1 of Amaya, Filbien, Okou and Roch, 2018).

**Assumption 2:** (Markov Trades). The trade direction indicator is the outcome of a first order Markov process defined by the transition matrix:

\[
P = \begin{bmatrix}
p_{11} & p_{12} \\
p_{21} & p_{22}
\end{bmatrix}
\]  

where the entries $p_{jk}$, $j, k = 1, 2$ are transition probabilities defined by: $\text{Prob}(q_t = k | q_{t-1} = j, q_{t-2} = l, \ldots) = \text{Prob}(q_t = k | q_{t-1} = j) = p_{jk}$ for $q_t = \pm 1$. The Markov trade direction indicators model the possibility of auto correlation of trades common in financial markets due to splitting of large orders to reduce market impact. It can also be used as an approximation of buying or selling pressure: $p_{11} > p_{22}$ meaning there are on average more buys than sells which should push prices up. By definition, the unconditional probabilities of buys, sells are: $\pi_1 = \frac{p_{21}}{p_{21} + p_{12}}, \pi_2 = \frac{p_{12}}{p_{21} + p_{12}}$, and the $k$-step transition probability matrix $P^k$ is given by:

\[
P^{(k)} = \begin{bmatrix}
\pi_1 & \pi_2 \\
\pi_1 & \pi_2
\end{bmatrix} + \lambda^k \begin{bmatrix}
\pi_2 & -\pi_2 \\
-\pi_1 & \pi_1
\end{bmatrix},
\]  

where $\lambda^k = (1 - p_{12} - p_{21})^k$ (Douc and Stoffer, 2014). This property will become important in the mid price update process and in defining the properties of the transactions prices.

**Assumption 3:** (Price Update Rule.) The mid price follows the process:

\[
m_{t+1} = m_t + s_{t+1}(q_t - \hat{q}_{t+1}) + u^m_{t+1}, \quad \text{where } \hat{q}_{t+1} = \mathbb{E}[q_{t+1} | q_t] \quad \text{and} \quad u^m_{t+1} \sim N(0, \sigma^2_m).
\]  

$\hat{q}_{t+1}$ is the prediction of the next trade sign given the sign of the last observed transaction and $u^m_t$ is an innovation to the mid price reflecting public information that is unrelated to the sequence of trade signs. Given the Markov property of trade signs, the best linear
2.2. Transactions history dependent prices formation

forecast for one step ahead sign is given by:

\[ \hat{q}_{t+1} = \mathbb{E}[q_{t+1}|q_t] = q_t \times \text{Prob}(q_{t+1} = q_t) - q_t \times \text{Prob}(q_{t+1} \neq q_t) = (1 - 2\pi)q_t, \]  

(2.5)

where \( \pi \) is the probability that the sign of a trade at \( t+1 \) is the opposite of the sign at \( t \):

\[ \pi = \text{Prob}(q_{t+1} \neq q_t) = \text{Prob}(q_{t+1} = +1, q_t = -1) + \text{Prob}(q_{t+1} = -1, q_t = +1) = 1 - (\pi_1 p_{11} + \pi_2 p_{22}). \]

Using conditional probability:

\[ \text{Prob}(q_{t+1} = +1, q_t = -1) = \text{Prob}(q_{t+1} = +1 \cap q_t = -1) \equiv P(A \cap B) = P(A|B)P(B) = \pi_1 \pi_2. \]

The trade related update of the mid-price change is then:

\[ s_t(q_t - \hat{q}_{t+1}) = 2\pi s_t q_t. \]  

If a buy(sell) is likely to be followed by another buy(sell) rather than a sell(buy), then \( p_{11} > \frac{1}{2} (p_{22} > \frac{1}{2}) \) and \( \pi_1, \pi_2 > 0 \) so that \( 0 < \pi < 1 \). It follows that if \( q_t = +1 \), then the efficient price moves up by a factor proportional to the spread times \( \pi \). For instance, if \( p_{11} = p_{22} = \frac{3}{4} \), then \( \pi_1 = \pi_2 = \frac{1}{2} \) and \( \pi = \frac{1}{4} \). The mid-price then moves up or down by the half-spread \( s_t \). If buys and sells are equally likely with \( \pi_1 = \pi_2 = 0.99 \) then is still \( \pi_1 = \pi_2 = \frac{1}{2} \) but \( \pi = 0.01 \) and the order flow has almost no effect on mid price movements which are then mainly driven by the non-trade related innovations \( u_t^m \).

In the absence of any changes to fundamental value, the quote adjustment in response to a transaction can be thought of as consisting of two components: (i) \( s_t q_t \), the private information revealed by the last trade and (ii) \( -s_t \hat{q}_{t+1} \), the compensation for inventory risk (Huang and Stoll, 1998). Since \( \pi \) is the probability of a trade sign reversal, the expected quote change, \( \mathbb{E}(\Delta m_{t+1}|m_t, q_t) = 2\pi s_t q_t \), measures how much of the inventory induced quote adjustment is expected to be reversed in the subsequent trade.

The three assumptions together lead to three properties of the price process: martingale prices, a quoted-spread and no quasi-arbitrage or price-manipulation. We discuss these properties in the next subsection as propositions. Proposition 1 shows how martingale prices result directly from the price-update rule of Assumption 2. Proposition 2 shows that the efficient price \( m_t \) in equation (2.1) coincides with quote mid-point, the average of the highest bid and lowest ask prices. Finally Proposition 3 shows that the price process rules out quasi–arbitrage.

2.2.2 Properties of the Price Process under Market Impact

Proposition 1. Martingale Prices: The transaction price process is a martingale.

Proof. The martingale property requires us to show: \( \mathbb{E}(p_{t+1}) = p_t \). We can write the trans-

\[ \end{document} \]
2. PRICE IMPACT

action price at time $t + 1$ as:

$$p_{t+1} = m_{t+1} + s_{t+1}q_{t+1} = m_t + s_{t+1}(q_t - \tilde{q}_{t+1}) + s_{t+1}q_{t+1} + u_{t+1}^m$$

$$= m_t + s_tq_t - s_tq_t + s_{t+1}(q_t - \tilde{q}_{t+1}) + s_{t+1}q_{t+1} + u_{t+1}^m$$

$$= p_t + q_t[s_{t+1} - s_t] + s_{t+1}[q_{t+1} - \tilde{q}_{t+1}] + u_{t+1}^m$$

and therefore find,

$$\mathbb{E}[p_{t+1}] = p_t + q_t \times \mathbb{E}[s_{t+1} - s_t] + \mathbb{E}[s_{t+1}] \times \mathbb{E}[q_{t+1} - \tilde{q}_{t+1}] + \mathbb{E}(u_{t+1}^m)$$

$$= p_t.$$  

The last line has used the assumptions $\mathbb{E}(s_{t+1}) = \mathbb{E}(s_t) = s$ and $\mathbb{E}(u_{t+1}^m) = 0$ together with the law of iterated expectations $\mathbb{E}(q_{t+1}) = \mathbb{E}(\mathbb{E}(q_{t+1}|q_t)) = \hat{q}_{t+1}$. We therefore conclude that

$$\mathbb{E}[p_{t+1}] = p_t. \quad \square$$

**Proposition 2. Bid-Ask Spread:** When agents set bid and ask prices so as not to have ex-post regrets, then the spread equals $2s$.

**Proof.** Given the sign forecast $\hat{q}_{t+1} = (1-2\pi)q_t$ defined in equation (2.5), we can write the efficient and transaction price processes as:

$$m_{t+1} = m_t + 2\pi s_{t+1}q_t + u_{t+1}^m, \quad \text{and} \quad p_{t+1} = m_t + (2\pi q_t + q_{t+1})s_{t+1} + u_{t+1}^m. \quad (2.6)$$

For either a buy and sell order, regret free price quotations require that the ask and bid prices are respectively set such that:

$$p_t^{\text{Ask}} = \mathbb{E}_t[p_{t+1}|q_{t+1} = +1] = m_t + (1 + 2\pi q_t)s \quad \& \quad p_t^{\text{Bid}} = \mathbb{E}_t[p_{t+1}|q_{t+1} = -1] = m_t - (1 - 2\pi q_t)s,$$

which implies the bid–ask spread given by: $p_t^{\text{Ask}} - p_t^{\text{Bid}} = 2s$. Furthermore, the mid-price before transaction occurring at time $t + 1$ coincides with the efficient price of the Roll model as:

$$\text{mid}_{t+1} = \frac{p_t^{\text{Ask}} + p_t^{\text{Bid}}}{2} = m_t + 2\pi q_t s \approx m_{t+1}$$

given by equation (2.4). \quad \square

**Proposition 3. No Quasi-Arbitrage:** The transaction price process $p_t$ does not admit quasi-arbitrage or price-manipulation of Huberman and Stanzl (2004).

**Proof.** Huberman and Stanzl (2004) define price dynamics for an order driven market where price updates and transaction prices coincide as: $p_t + 1 = p_t + s_{t+1}(Q_{t+1} + \eta_{t+1}) + u_{t+1}$, where the real number $s_{t+1}$ measures the liquidity with respect to trading volume or order size $Q_{t+1}$, $\eta_{t+1}$ is a random variable representing the unknown volume of the other market participants and $u_{t+1}$ incorporates non-trade related news into the price process. No arbitrage or price manipulation requires that the stochastic processes $\eta_{t+1}$ and $u_{t+1}$ be zero mean i.i.d random variables with constant variances.
2.2. Transactions history dependent prices formation

From the martingale property, we can write the transaction price as:

\[ p_{t+1} = p_t + q_t \left( s_{t+1} - s_t \right) + s_{t+1} \left( q_{t+1} - \hat{q}_{t+1} \right) + u_{t+1} \]

\[ = p_t + s_{t+1} \left( q_t \left( \frac{s_{t+1} - s_t}{s_{t+1}} + \frac{q_{t+1} - \hat{q}_{t+1}}{\eta_{t+1}} \right) + u_{t+1} \right) \]

\[ \equiv p_t + s_{t+1} \left( Q_{t+1} + \eta_{t+1} \right) + u_{t+1} \]

which is equivalent to Huberman and Stanzl’s (2004) equation with appropriate reinterpretation of variables: \( Q_{t+1} \) is the order size, which now scales with the real number \( 1 - s_t / s_{t+1} \) and \( \eta_{t+1} \) is the random variable representing the demand of other market participants. The only condition we need to rule out price manipulation is that \( \eta_{t+1} \) be an i.i.d. stochastic process with zero mean and constant variance:

\[ \mathbb{E}[\eta_{t+1}] = 0 \quad \text{and} \quad \text{Var}[\eta_{t+1}] = \mathbb{E}(\eta_{t+1} - \mathbb{E}[\eta_{t+1}])^2 = \mathbb{E}(q_{t+1} - \hat{q}_{t+1})^2 = \text{Var}[q_{t+1}] = 4\pi_1\pi_2. \]

This completes the proof of the claim.

\[ \square \]

2.2.3 Empirical Predictions

We now consider the empirical predictions that derive from our assumptions about the sign sequences generation and the price update process. The price update process gives two theoretical predictions that can be tested in the data. We define the unconditional impact and response functions followed by stochastic volatility.

**Unconditional Impact and Response Functions** Recalling that \( m_t \) denotes the mid-price immediately before the transaction at time \( t \), the lag-1 unconditional impact is defined as (Bouchaud et al., 2006):

\[ R(1) := \langle (m_{t+1} - m_t) \cdot q_t \rangle_t, \]

where the empirical average \( \langle \cdot \rangle_t \) is taken over all transactions of any volume. For any \( k > 0 \), we can define the response function:

\[ R(k) = \mathbb{E}[(m_{t+k} - m_t) \cdot q_t] \equiv \langle (m_{t+k} - m_t) \cdot q_t \rangle_t, \quad (2.7) \]

which measures the information content of the current trade on the mid-price \( k \) trades into the future. Our assumptions about the price update process equation (2.4) means that the lag-\( k \) forward difference can be written as the telescoping sum:

\[ m_{t+k} - m_t = \sum_{\ell=1}^{k} (m_{t+\ell} - m_{t+\ell-1}) = \sum_{\ell=1}^{k} s_{t+\ell} (q_{t+\ell} - \hat{q}_{t+\ell}) + \sum_{\ell=1}^{k} u_{t+\ell} \]

\[ = 2\pi \cdot \sum_{\ell=1}^{k} s_{t+\ell} q_{t+\ell-1} + u_{t+k}. \]
2. PRICE IMPACT

Using this relation and noting that conditioning on \( q_t \), we can use the law of iterated expectations to compute \( \mathbb{E}[q_{t+k} q_t] \) as:

\[
\mathbb{E}(q_{t+k} q_t) = \mathbb{E}_t(\mathbb{E}(q_{t+k} | q_t)) = \pi_1 \cdot \mathbb{E}(q_{t+k} | q_t = +1) - \pi_2 \cdot \mathbb{E}(q_{t+k} | q_t = -1) = a + b \lambda^k,
\]

where \( a = (\pi_1 - \pi_2)^2 \), \( b = 4 \pi_1 \pi_2 \) and \( \lambda = 1 - p_{12} - p_{21} \) follows from equation (2.3); we can express the response function as:

\[
R(k) = \mathbb{E}[(m_{t+k} - m_t) \cdot q_t] = 2\pi \sum_{\ell=1}^{k} \mathbb{E}(s_{t+\ell} \times q_{t+k-1} q_t)
= 2\pi s \cdot \left(1 + [a + b \lambda] + \cdots + [a + b \lambda^{k-1}] \right)
= 2\pi s \cdot \left(1 - b + a(k-1) + b \frac{1 - \lambda^k}{1 - \lambda} \right).
\]

Equation (2.8) says that the response function contains the reaction impact of future trades adjusted to their correlation with the present trade. Defining the lag-\( k \) anti-correlation function:

\[
C(k) = a(k-1) - b \left(1 - \frac{1 - \lambda^k}{1 - \lambda} \right), \quad \text{for } k > 0, \text{ with } C(1) = 0,
\]

we find the following one-to-one relationships between lag-1 and lag-\( k \) response functions:

\[
R(1) = \frac{1}{1 + C(k)} \cdot R(k) \equiv R(k+n) = \frac{1 + C(k+n)}{1 + C(k)} \cdot R(k), \text{ for } n, k = 1, 2, \ldots
\]

The empirical relationship (2.10) does not appear in the original MRR model (even though that model is Markovian as well). Similar expressions were first derived by Wyart, Bouchaud, Kockelkoren, Potters and Vettorazzo (2006), but with \( R(k) = \frac{1 - C(k)}{1 - C(1)} R(1) \). This different version of the response function follows from the MRR assumption of sign surprises moving the mid-price (Bonart and Lillo, 2018).

Volatility and the Spread Stochastic volatility over \( k \) trades is defined by the average:

\[
\frac{1}{k} \sum_{\ell=1}^{k} [\Delta p_{t+\ell}]^2.
\]

Approximating trade by trade price change by the mid-price change and using equation (2.6), the price difference between any two trades is \( \Delta p_{t+1} \approx \Delta m_{t+1} \approx 2\pi s_{t+1} q_t \) and we can approximate volatility over \( k \) trades by the empirical average:

\[
\frac{1}{k} \sum_{\ell=1}^{k} \mathbb{E}[\Delta p_{t+\ell}]^2 \approx \langle 4\pi^2 \times (s_t^2 + \sigma_s^2) \rangle_k.
\]

Equation 2.11 gives the fraction of price volatility that can be attributed to trading and/or microstructure effects as it excludes any fundamental information components \( u_{t+1}^m \).
2.2.4 Relation to Literature

Informed trading due to the arrival of new information that permanently moves prices has been studied in the literature as the price impact of a trade (Biais, Glosten and Spatt, 2005). The ideas of price impact (permanent and temporary) have been found in the literature since at least the work of Almgren and Chriss (2001) and appeared in many other studies including Ponzi, Lillo and Mantegna (2009); Obizhaeva and Wang (2013); Sun, Kruse and Yu (2013) among others. The idea that permanent impact is the result of a surprise in the order flow originates from the MRR model (Madhavan, Richardson and Roomans, 1997). In the MRR model, the efficient price evolves because of random external shocks or “news” and trade impact. Both trade sizes and the measure of a trade’s impact are assumed constant but the direction indicators are correlated and therefore a trade has long time impact (Wyart et al., 2006; Mizrach and Otsubo, 2014).

Market impact has also been studied within the context of the propagator model introduced by Bouchar, Kockelkoren and Potters (2006). Taranto, Bormetti and Lillo (2014) model the sign predictor of equation (2.5) as a discrete auto-regressive process of order $k$ – DAR($k$) – and derive linear impact functions. These models usually involve a mid-price update that depends on the mismatch between the sign of the present trade and what the market marker predicted rather than the difference between the current trade sign and the prediction of the next trade as we assume. Mid-price updates related to the spread have also been considered by Curato and Lillo (2015) who design models in which the “return process depends on the transition between two subsequent spread states, distinguishing the case in which the spread remains constant and the case when it changes” (pg. 6). Tóth, Eisler and Bouchaud (2017) have recently found that the market impact of a trade is enhanced by other copy-cat trades but the effect subsequently decays fast. The linearity of impact functions originates from the work of Glosten and Milgrom (1985); Kyle (1985) and can be found in the works of Obizhaeva and Wang (2013) among others. Huberman and Stanzl (2004) and Gillemot, Farmer and Lillo (2006) show that only linear impact functions exclude the possibility of price manipulation. This is formalized in Proposition 3 where it is shown that our assumptions exclude quasi-arbitrage and price manipulation.

In financial econometrics, it has been known for a long time that microstructure noise, including bid-ask spreads, induce large biases in estimates of integrated variance (Bandi and Russell, 2006; Barndorff-Nielsen, Hansen, Lunde and Shephard, 2009). Zhang, Mykland and Aït-Sahalia (2005) show that in the presence of microstructure noise, the standard estimate of realized volatility simply captures the variance of the noise. Aït-Sahalia, Mykland and Zhang (2005) have shown that it is necessary to model microstructure noise and use all data available at a high frequency, rather than the arbitrary sampling windows

\[ \Delta p_t = \Delta p_{t+1} = G_0(1)q_t + \sum_{i>0}[G_0(i+1) - G_0(i)]q_{t-i} + \eta_t. \]
generally adopted by practitioners, while Hansen and Lunde (2006); Hansen, Large and Lunde (2008) have developed kernel based methods for consistent estimation of realized variance in the presence of transaction induced noise such as the bid-ask spread. In testing empirical predictions in Sec. 2.4, we find that a large fraction of the volatility over 15 minute intervals can be explained by the spread and the probability of a trade sign reversal. This is similar to Plerou et al. (2005) who show that the bid-ask spread and volatility are also related logarithmically; and to Benzaquen et al. (2017) finding that for futures markets, volatility is proportional to the spread and the square root of the number of trades above a given threshold.

2.3 Calibration

In order to test the empirical predictions of the model, we need to estimate the parameters of the Markov trade signs, the spread and the mid-prices. The standard approach is usually to obtain high frequency data from services such as LOBSTER\textsuperscript{18} which provides up to millisecond time stamped data sets for stocks, with quotes, signs and volumes. These have been used by Curato and Lillo (2014, 2015); Bonart and Lillo (2018) among others. Our data come from the Tokyo Commodity Exchange (TOCOM)\textsuperscript{19}. The data from TOCOM are trade-by-trade with time stamps, but come without information about the initiator side of each transaction (even though five minute summaries of quote updates are available). A full description of the data is given in Sec. 2.3.2. To obtain the parameters of interest, we estimate a state space system as described below.

2.3.1 State Space Representations and Estimation

We work with differences of observed tick-by-tick transaction prices and write the price change equation in the state-space form:

$$y_t \equiv \Delta p_t = m_t - m_{t-1} + s_t q_t - s_{t-1} q_{t-1} = A_t x_t,$$  \hspace{1cm} (2.12)

where $x_t = (m_t, m_{t-1}, s_t, s_{t-1})'$ defines a (4 x 1) unobserved state vector that follows the first autoregressive process $x_t = \Phi x_{t-1} + Y + u_t$:

$$
\begin{pmatrix}
    m_t \\
    m_{t-1} \\
    s_t \\
    s_{t-1}
\end{pmatrix} = 
\begin{pmatrix}
    1 & 0 & 0 & 0 \\
    1 & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 \\
    0 & 0 & 1 & 0
\end{pmatrix} 
\begin{pmatrix}
    m_{t-1} \\
    m_{t-2} \\
    s_{t-1} \\
    s_{t-2}
\end{pmatrix} +
\begin{pmatrix}
    0 \\
    0 \\
    0 \\
    0
\end{pmatrix} +
\begin{pmatrix}
    \hat{u}_m^t \\
    0 \\
    \hat{u}_s^t \\
    0
\end{pmatrix},
$$  \hspace{1cm} (2.13)

where $A_t = \begin{bmatrix} 1, -1, q_t, -q_{t-1} \end{bmatrix}$ is a (1 x 4) observation or measurement matrix, and the error vector $u_t = (\hat{u}_m^t, 0, \hat{u}_s^t, 0)'$ is i.i.d $N(0, \Sigma_u)$ where $\Sigma_u$ is a (4 x 4) variance–covariance matrix

\textsuperscript{18}https://lobsterdata.com/
\textsuperscript{19}https://www.tocom.or.jp/historical/download.html
with off-diagonal elements equal to zero and diagonal \((\hat{\sigma}_m^2, \sigma_s^2, 0)\). The observation matrix \(A_t\) is a first order Markov process that takes the values:

\[
A_t = \{ \tilde{A}_1, \tilde{A}_2, \tilde{A}_3, \tilde{A}_4 \} = \{ [+1, -1, +1, -1], [+1, +1, +1, +1], [+1, -1, -1, -1], [+1, -1, -1, +1] \},
\]

(2.14)

with transition probability matrix \(\tilde{P}\) whose entries are defined by: \(\tilde{p}_{jk} = \text{Prob}(A_t = \tilde{A}_k | A_{t-1} = \tilde{A}_j, A_{t-2} = \tilde{A}_l, \ldots) = \text{Prob}(A_t = \tilde{A}_k | A_{t-1} = \tilde{A}_j)\) for \(j, k, l \in \{A_t\}:

\[
\tilde{P} = \begin{pmatrix}
p_{11} & 0 & p_{12} & 0 \\
p_{11} & 0 & p_{12} & 0 \\
0 & p_{21} & 0 & p_{22} \\
0 & p_{21} & 0 & p_{22}
\end{pmatrix}
\]

(2.15)

where the \(p_{jk}\) are given by equation (2.3) and preserve the normalization \(\sum_{k=1}^{4} \tilde{p}_{jk} = 1\). The entries of this \(4 \times 4\) matrix follow directly from the relationship between \(A_t\) and the vector \(\tilde{q}_t = (q_t, q_{t-1})\).

The parameters to be estimated are \(\Theta_1 = (s, \sigma^m, \sigma^s)\) and \(\Theta_2 = (p_1, p_2)\); the last two defining the transition matrix probabilities \(p_{kk} = \exp \alpha\), \(k = (1,2)\). Letting \(\epsilon_{t,\ell} = (y_t - \tilde{A}_{\ell} x_t^{-1})\) denote the innovation from equation (2.12) when \(A_t = \tilde{A}_{\ell}, \ell = 1, 2, 3, 4\), then the log-likelihood is given by:

\[
\ln L(y_{1:T}; \Theta_1, \Theta_2) = \sum_{t=1}^{T} \ln \left( \sum_{\ell=1}^{4} f(\epsilon_{t,\ell}; \Theta_1) \times \pi_{\ell}(t | t-1; \Theta_2) \right)
\]

where \(f(\epsilon_{t,\ell}; \Theta_1)\) is the normal density with mean 0 and variance \(\Sigma_{\ell} \); with \(\pi_{\ell}(t | t-1; \Theta_2) = \text{Prob}[A_t = \tilde{A}_{\ell} | \psi_{t-1}]\) being the probability of the measurement matrix taking particular values given by equation (2.14). Evaluating the likelihood requires computation of the predicted probabilities \(\pi_{\ell}(t | t-1; \Theta_2)\) which in turn require computing the filtered probabilities \(\pi_{\ell}(t | t; \Theta_2, \psi_t)\) after observing the data \(y_t = \Delta p_t\).

Collecting the predicted and filtered probabilities into \((4 \times 1)\) vectors \(\xi_{t-1}^{\ell}\) and \(\xi_{t}^{\ell}\), respectively, and the respective densities for each \(\ell\) at time \(t\) in the \((4 \times 1)\) vector \(\eta_t = f(\epsilon_{t,\ell}; \Theta_1 | \tilde{A}_{\ell}, \psi_{t-1})\), the updates are accomplished as follows:

\[
\xi_{t|t-1}^{\ell} = \xi_{t-1|t-1}^{\ell} \tilde{P}, \quad \xi_{t|t}^{\ell} = \frac{\xi_{t|t-1}^{\ell} \odot \eta_t}{\sum \xi_{t|t-1}^{\ell} \odot \eta_t}
\]

(2.16)

where \(\tilde{P}\) defined by (2.15), \(\odot\) is the Hadamard product and the summation is over 4 states.

I now establish the recursions for the filters associated with the state \(x_t\); with those for the switching process \(A_t\) remaining as defined in (2.16). The predictors \(x_{t}^{\ell-1} = E(x_t | y_{1:t-1})\) and filters \(x_{t}^{\ell} = E(x_t | y_{1:t})\), and their associated error variance–covariance matrices, \(P_{t}^{\ell-1} =

\text{Note that this specification does not include the price update part for } m_t \text{ so our error is } \tilde{u}_t^m = u_t^m + s_t(\tilde{q}_{t-1} - \tilde{q}_t) \text{ with } E[\tilde{u}_t^m] = 2\pi s x (\pi_1 - \pi_2) \approx 0 \text{ and } \text{Cov}(\tilde{u}_t^m, u_t^f) = 0.
2. PRICE IMPACT

\[
E\{ (x_t - x_{t-1}^{(t)}) (x_t - x_{t-1}^{(t)})' \} \text{ and } P_t = \text{Cov} \{ x_t | y_{1:t-1}, \epsilon_t \} \text{ respectively, are given by:}
\]

\[
x_t^{(t)} = \Phi x_{t-1}^{(t-1)}, \quad x_t = x_t^{(t)} + \sum_{\ell=1}^{4} \pi_{\ell}(t) K_{t,\ell} \left( y_t - \bar{A}_{\ell} x_t^{(t-1)} \right) \quad (2.17)
\]

\[
P_t^{(t)} = \Phi P_{t-1}^{(t-1)} \Phi' + \Sigma_u, \quad P_{t|t} = \sum_{\ell=1}^{4} \pi_{\ell}(t) \left( I - K_{t,\ell} \bar{A}_{\ell} \right) P_t^{(t)} \quad (2.18)
\]

where \( K_{t,\ell} = P_t^{(t-1)} \bar{A}_{\ell} \Sigma_t^{-1} \) is the Kalman gain and \( \Sigma_{t,\ell} = \text{Var}(\epsilon_{t,\ell}) = \bar{A}_{\ell} P_t^{(t-1)} \bar{A}_{\ell}' \) is the variance of the innovation \( \epsilon_{t,\ell} = (y_t - \bar{A}_{\ell} x_t^{(t-1)}) \). Equations (2.17) and (2.18) exhibit the filter values as weighted linear combinations of the innovation values \( \epsilon_{t,\ell} \) and \( \Sigma_{t,\ell} \) corresponding to the four measurement matrices \( \bar{A}_{\ell}, \ell = 1, 2, 3, 4 \). These equations are similar to the approximations first introduced by Bar-Shalom and Tse (1975); Masreliez and Martin (1977) and Peña and Guttman (1989).

2.3.2 Data Description

Our data consists of two of the most liquid commodity futures contracts traded at the Tokyo Commodity Exchange: Gold Standard (TOCOM Product Code: 11, Reuters Contract Detail: TCE / JAU, Bloomberg Ticker: JGA ) and Platinum Standard (TOCOM Product Code: 13, Reuters Contract Detail: TCE / JPL, Bloomberg Ticker: JAA). The contracts are for physical delivery of 1kg/contract (approximately 32.15 troy ounces) of Gold and 500g/contract (approximately 16.08 troy ounces) of Platinum.\(^{21}\) Each of the contracts has a minimum price increment of JPY 1 per gram.

The data cover the day-time trading session, from 8:45 a.m. to 3:15 p.m. Japanese Standard Time (JST) on the 24th April 2019, with the delivery month of February 2020. There are two trading sessions per 24 hour day. The Day Session with an opening call auction (Ita-awase) at 8:45 a.m. (JST) followed by continuous trading (Zaraba) till 3:10 p.m. (JST) and a closing call auction (Ita-awase) at 3:15 p.m. (JST). The Night Session operates in a similar fashion but with the opening auction at 4:30 p.m. (JST) followed by continuous trading till 5:25 a.m. of the following morning and a closing auction at 5:30 a.m. at the end of continuous trading. The terms Ita-awase and Zaraba refer to the methods used to match buyers and sellers. The Ita-awase method requires three conditions to be met at the call auction: (i) all market orders, sell or buy orders, must be executed, (ii) all limit orders to sell/buy at a price higher/lower than the execution price must be executed and, (iii) at the execution price, the entire amounts of either all sell or all buy orders, and at least one trading unit from the opposite side of the order book must be executed (Taniguchi, Ono and Mori, 2008; Aruka, 2015). After the opening auction, the Zaraba method is used to determine prices: a new order results into a transaction if the price matches an order that has already been received, with price and time priority applied as appropriate.\(^{22}\)

\(^{21}\)See https://www.tocom.or.jp/products/index.html for detailed descriptions of these and other contracts traded at the exchange.

\(^{22}\)Time priority: for buy orders, the order at a high price has priority over the order at a low price, while the order at a low price has priority over the order at a high price for sell orders. Price priority: when orders are
2.3. Calibration

We analyze data for the day session on 24th April 2019 (the market is more active during the day session). Over the six and half hour period beginning at 8:45 to 15:15, there are 1,633 and 1,566 trades of the Gold and Platinum contracts respectively. Table 2.1 gives a summary of the descriptive statistics of the commodities and Figures 2.2a, 2.2b show the time series of prices every 5 minutes.

![Figure 2.2: Open, Close, Min and Max Transaction Prices, 24th April 2019](image)

(a) TCE/JAU – Gold

(b) TCE/JPL – Platinum

Figures 2.3a and 2.3b show the tick-by-tick distribution of returns and traded volume. The Gold returns are almost always zero (80%), with the return +1 and -1 making about 10% each. Only one return equals +2, while three returns equal to -2. Platinum returns are more spread: 70% at zero, $\pm 1$ at about 13% each and $\pm 2$ at 1% each and 3 returns equal to -3 and one at 3. The returns are discrete values since the minimum price increment in each case equal 1 ¥. We will hence model the data in their natural units rather than log-returns. Most trades are of a single contract (53% for Gold and 64% for Platinum) and about 14% of trades equal two contracts.

placed at the same price, an earlier order takes precedence over subsequent orders (Taniguchi, Ono and Mori, 2008).
2. PRICE IMPACT

Table 2.1: Summary Statistics of Prices over 5 minute Intervals

<table>
<thead>
<tr>
<th>Contract</th>
<th>Symbol</th>
<th>T</th>
<th>Date</th>
<th>Prices in Japanese Yen (¥)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOLD</td>
<td>11</td>
<td>1633</td>
<td>24-Apr-2019</td>
<td>Mean  Min  Max  Std. Dev.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Close   4558.78 4553 4565  3.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Open    4558.73 4553 4565  3.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High    4559.51 4554 4566  3.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low     4558.10 4553 4564  3.51</td>
</tr>
<tr>
<td>PLATINUM</td>
<td>13</td>
<td>1566</td>
<td>24-Apr-2019</td>
<td>Mean  Min  Max  Std. Dev.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Close   3189.31 3182 3197  4.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Open    3189.44 3182 3197  4.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High    3190.30 3183 3197  4.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low     3188.31 3181 3196  4.29</td>
</tr>
</tbody>
</table>

Figure 2.3: Distribution of Returns and Traded Volumes

(a) Returns per Trade

(b) Volume per Trade

2.3.3 Estimation Results

Table 2.2 gives parameter estimates for the state space model described in subsection 2.3.1. We use the `fminunc` optimizer in MATLAB to perform the estimation. The initial values are set at half a tick for the spread and assuming buys and sells are approximately equal. All estimates are statistically significant and the estimation converges very fast. For this particular day session’s data, buyer initiated trades are more likely than seller initiated trades, with a buy(sell) likely to be followed by another buy(resp. sell) in both commodity contracts. For the Gold contract, the transition probability estimates $p_{11} = 0.8843, p_{22} = 0.7182$ imply that the unconditional probabilities of a buyer and seller initiated trades are approximately equal: $\pi_1 = \text{Prob}(q_1 = +1) = \frac{p_{11}}{p_{12} + p_{21}} = \frac{1 - 0.7182}{(1 - 0.8843) + (1 - 0.7182)} = 0.71$, $\pi_2 = 0.29$, so approximately 70% of trades are buyer initiated. The unconditional probability of a
2.3. Calibration

Sign reversal is \( \pi = 1 - (\pi_1 p_{11} + \pi_2 p_{22}) = 0.1655 \), meaning approximately 16% of transactions are followed by a transaction of the opposite sign. Given these estimates, we should expect that a buyer initiated trade results into a price update equal to: \( 2\pi s_t = 0.16 \text{ ¥} \) on average. The Platinum contract behaves similarly.

The spread is almost always equal to half a tick and the volatility of the spread component of returns is larger than that of non-trade related information: in each case \( \sigma_s > \sigma_m \).

This result is in line with Bandi and Russell (2006); Barndorff-Nielsen et al. (2009) who have showed that in the presence of microstructure noise, most standard volatility estimators simply capture this noise. We show in subsection 2.4 that most of the intra-day volatility is explained by the spread and sign auto-correlation.

| Table 2.2: Parameter Estimates of Switching Dynamic Linear Model |
|------------------|------------------|
| **Contract**     | **(1)**          | **(2)**          |
| **Parameter**    | **Gold**         | **Platinum**     |
|                  | Estimate S.E.    | Estimate S.E.    |
| \( s \)          | 0.4592 0.0181    | 0.5299 0.0113    |
| \( \sigma_m \)   | 0.1271 0.0115    | 0.1363 0.0145    |
| \( \sigma_s \)   | 0.1576 0.0139    | 0.2738 0.0132    |
| \( p_{11} \)     | 0.8843 0.0145    | 0.8319 0.0137    |
| \( p_{22} \)     | 0.7182 0.0254    | 0.6704 0.0284    |
| Log L            | -993.7987        | -245.4847        |
| \( T \)          | 1632             | 1565             |

**Bid-Ask Spreads** We then use the model to get filtered state estimates values of mid-prices, spreads and the probabilities for \( q_t = \pm 1 \). Figure 2.4 shows the variation of the effective bid-ask spread estimates. The spread is almost always equal to 1 ¥ (the minimum tick value), though occasionally rises up to 1 \( \frac{1}{2} \) ¥. This is a characteristic of a highly liquid market. Following the definition of Dayri and Rosenbaum (2015), we can consider these commodities as large-tick assets since the spread is almost always equal to 1 tick. These estimates are not corrected for the auto-correlation of order flow. When such a correction is made, the spread is almost always equal to \( \frac{1}{2} \) ¥. We show how to compute this correction in Appendix A.
2. PRICE IMPACT

2.4 Empirical Tests

We now perform the empirical tests of the predictions made in Sec. 2.2.3. We begin by comparing response functions from lag-1 to lag-4 [Equation (2.8)], then compare the model implied volatility to the actual values [Equation (2.11)].

Figures 2.5a and 2.5b shows the lag-1 empirical test of the prediction (2.8) for Gold and Platinum respectively. Each point represents the mean of $R(1) = \frac{1}{1+C_{k}} \cdot R(k)$ over 60 trades (approximate number of transactions per 15 minute interval: there are 6.5 hours between 8:45-15:15 and 1633(1566) trades over the session, which gives $\frac{1633}{(6.5 \times 4)} \approx 60$ trades every 15 minutes). These show that our hypothesis about the price update rule used by agents is on average correct. This is quite remarkable given the averages are over the relatively small samples of 60 trades. This relationship remains when the averaging is performed over larger samples up to 150 trades or over time intervals from 5 to 30 minutes. To find a similar relationship, Bonart and Lillo (2018) average returns during the whole year 2015 for NASDAQ stocks.

Figures 2.6a and 2.6b shows the lag-$k$ empirical test of the prediction (2.8) for Gold and Platinum Futures Prices. Each cloud of points represents the mean of $R(k+1) = \frac{1+C_{k+1}}{1+C_{k}} \cdot R(k)$ averaged over 60 trades. The model predictions are matched almost perfectly: the intercepts are statistically not different from zero and the slopes are almost equal to unity.

Table 2.3 shows simple OLS regression tests of the two versions of equation (2.8). In the first case, where we run the regression $R(1) = \alpha + \beta \frac{1}{1+C_{k}} \cdot R(k)$. If our hypothesis is correct, then $\alpha$ should equal zero and $\beta$ unity. The results of this regression are shown in the first four rows (standard errors in parenthesis). In all cases $\alpha = 0$ and $\beta = 1$, with $R^{2}$ values of at least 50%. The last three rows show the results of the regression $R(k+1) = \alpha + \beta \frac{1+C_{k+1}}{1+C_{k}} \cdot R(k)$.
the alternative version of equation (2.8). Again the coefficients \( \alpha \) and \( \beta \) are of the expected magnitude, zero and unity respectively. This version of the test gives \( R^2 \) values at least 90\%. These results support our hypothesis about the nature of mid-price updates for these commodity futures prices.

**Table 2.3:** Regression Tests of Relation between Lags-1 to \( k \) and Lags-\( k \) to \( k + 1 \).

<table>
<thead>
<tr>
<th>( y = \alpha + \beta x )</th>
<th>( x )</th>
<th>( y )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \text{Adj.} R^2 )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \text{Adj.} R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R(1) ) ( \frac{1}{1+C_{(2)}} ) ( R(2) )</td>
<td>0.0004</td>
<td>1.0185</td>
<td>0.8976</td>
<td>0.0021</td>
<td>1.0189</td>
<td>0.7468</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0013)</td>
<td>(0.0700)</td>
<td>(0.0032)</td>
<td>(0.1179)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R(1) ) ( \frac{1}{1+C_{(3)}} ) ( R(3) )</td>
<td>-0.0006</td>
<td>1.0937</td>
<td>0.8906</td>
<td>0.0072</td>
<td>0.8852</td>
<td>0.5657</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0014)</td>
<td>(0.0781)</td>
<td>(0.0039)</td>
<td>(0.1528)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R(1) ) ( \frac{1}{1+C_{(4)}} ) ( R(4) )</td>
<td>-0.0001</td>
<td>1.1368</td>
<td>0.7659</td>
<td>0.0100</td>
<td>0.8629</td>
<td>0.4477</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0021)</td>
<td>(0.1275)</td>
<td>(0.0042)</td>
<td>(0.1871)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R(1) ) ( \frac{1}{1+C_{(5)}} ) ( R(5) )</td>
<td>0.0007</td>
<td>1.1090</td>
<td>0.6601</td>
<td>0.0110</td>
<td>0.9221</td>
<td>0.5030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0025)</td>
<td>(0.1607)</td>
<td>(0.0037)</td>
<td>(0.1798)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R(3) ) ( \frac{1+C_{(2)}}{1+C_{(2)}} ) ( R(2) )</td>
<td>0.0032</td>
<td>0.9223</td>
<td>0.9879</td>
<td>-0.0054</td>
<td>0.9581</td>
<td>0.9060</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0012)</td>
<td>(0.0208)</td>
<td>(0.0049)</td>
<td>(0.0616)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R(4) ) ( \frac{1+C_{(3)}}{1+C_{(3)}} ) ( R(3) )</td>
<td>0.0025</td>
<td>0.8701</td>
<td>0.9417</td>
<td>-0.0050</td>
<td>0.9191</td>
<td>0.9695</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0032)</td>
<td>(0.0441)</td>
<td>(0.0032)</td>
<td>(0.0326)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R(5) ) ( \frac{1+C_{(4)}}{1+C_{(4)}} ) ( R(4) )</td>
<td>0.0012</td>
<td>0.9418</td>
<td>0.9700</td>
<td>-0.0072</td>
<td>0.9623</td>
<td>0.9077</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0028)</td>
<td>(0.0338)</td>
<td>(0.0067)</td>
<td>(0.0612)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.7 shows the relationship between the model implied versus actual volatility as shown in the relation given by equation (2.11). Again, the model fits almost perfectly: the intercepts are statistically not different from zero, the slopes equal unity and \( R^2 \geq 0.85 \). This implies that most of the price volatility can be accounted for by variations in the spread and sign autocorrelation as hypothesized. The intercepts being zero means that there is almost no news component in the session’s price fluctuations. (Wyart et al., 2006) make a similar observation in the case of stocks and explain this observation to be due to the volatility of stocks being “mostly attributed to market activity and trade impact”.
2. PRICE IMPACT

Figure 2.5: Lag-1 to $k$ Response Functions

Notes: Each cloud of points, consisting of a diamond (rescaled $R(2)$), a plus (rescaled $R(3)$) and a circle (rescaled $R(4)$), corresponds to a non-overlapping average every 60 trades through the day session. The dotted line ($y = x$) is a guide to the eye and the gray line is linear fit $y = \alpha + \beta x$ shown in the first three rows of Table 2.3.

Figure 2.6: Lag-$k$ to $k+1$ Response Functions

Each cloud of points, consisting of a diamond ($R(3)$ vs. rescaled $R(2)$), a plus ($R(4)$ vs. rescaled $R(3)$) and a circle ($R(5)$ vs. rescaled $R(4)$), correspond to non-overlapping averages over 60 trades through the day session. The dotted line ($y = x$) is a guide to the eye and the gray line is linear fit $y = \alpha + \beta x$ shown in the last 3 rows of Table 2.3.
2.4. Empirical Tests

**Figure 2.7: Volatility and the Spread**

(a) **GOLD**

Each circle corresponds to non-overlapping 15 minute averages through the day session. The dotted line \((y = x)\) is a guide to the eye and the gray line is linear fit: \(y = 0.0255 + 1.0096x, R^2 = 0.9311\).

(b) **PLATINUM**

Each circle corresponds to non-overlapping 15 minute averages through the day session. The dotted line \((y = x)\) is a guide to the eye and the gray line is linear fit: \(y = 0.0031 + 1.2536x, R^2 = 0.8517\).
2.5 Conclusion

We started this paper by postulating a theory of how market participants make price updates in the limit order book in response to the observed sequence of trades. Our hypothesis has been that agents actively trading in the market update their quotes on the basis of a prediction of the next trade sign. This “forecast” of the next trade sign is the best linear predictor made after observing the latest trade sign and the recent history of signs. Together with the assumptions about the signs generating process, the price update rule makes falsifiable empirical predictions about the impact response function and the volatility per trade. Our empirical tests suggest that the update rule with hypothesized and in agreement with empirical data.

Our aim has been to show that for commodity markets, the trading process itself leads to price discovery as it reveals the expected future value of the underlying physical commodity to be delivered at a later date. The fact that trades have price impact means that this intuition is correct: on average, an order flow imbalance in favor of buyer initiated trades leads to an uptick in prices and vice-versa.
CHAPTER 3

A Double Mixture Autoregressive Model of Commodity Prices

Abstract

Many commodity prices exhibit boom–bust type behavior: sustained periods of price increases are followed by sudden sharp collapses. Since around the year 2000, booms have become longer while busts have tended to be short but steep, suggesting a structural change in the behavior of prices. We model these features of the data using a novel double mixture autoregressive regression with two independent hidden Markov chains that characterize the data generating process. One chain models shifts in mean growth rates that accounts for rising and falling prices, while a second chain models changes in volatility and lag-structure. While the two chains are independent, the persistence of price growth depends of the volatility state, which allows autoregressive terms to vary across variance regimes. In order to maintain the model’s empirical relevance in dating price booms and busts while also correctly identifying lag–structure, a two–stage Fisherian approach to estimation is required. In the first stage, location related parameters are estimated while suppressing the underlying autoregressive terms of the series. In a second stage, the location related parameters are held fixed while determining the optimal number of autoregressive coefficients across the variance regimes. We apply the model to three industrial commodities price time series: Crude Oil, Aluminium and Rubber. We find that in each case, the model captures boom and bust cycles, with data from more recent periods exhibiting higher volatility, longer price rallies and steeper collapses.

Keywords: commodity price booms, hidden Markov models, regime switching, nuisance parameter problem, profile likelihoods, filtering.

JEL Codes: C5, C51, C58, C32, E32
3.1 Introduction

It is well known that commodity prices are subject to booms and busts over long time durations (Arezki, Hadri, Loungani and Rao, 2014). The boom phases are characterized by an increase in prices and low volatility while in the bust cycles, prices fall and volatility increases (Cashin, McDermott and Scott, 2002). These features of commodity price movements are similar to those of business cycles, where economic expansions are characterized by low volatility in macroeconomic time series with contractions being more volatile. One way to model time series subject to such cycles is the hidden Markov or regime switching model popularized in economics by Hamilton (1990). However, the two-state regime switching model is rarely used to date booms and busts in commodity prices. One potential reason is that the model breaks down when there are shifts in volatility that are independent from the boom-bust cycles, a characteristic of many commodity price time series. We present a model that can account for these volatility switches and successfully date the price cycles.

While this work analyzes commodity price data, its methods are inspired by developments in macroeconomic time series modeling of the business cycle. Recent implementations of Hamilton’s (1989) model have introduced a second Markov chain in regime switching models to capture volatility shifts (Bai and Wang, 2011; Doornik, 2013; Chauvet and Su, 2014). This change in the modeling of economic indicators such as GDP and unemployment was introduced after researchers noticed that single chain models performed poorly when the data were extended to include periods beyond 1984. Macroeconomic time series covering the period beginning from the second quarter of 1984 are generally characterized by low volatility, the so called Great Moderation, in multiple measures of economic activity. In order to maintain reasonable dating performance using Hamilton’s model, the Great Moderation related permanent shift in volatility has to be explicitly modeled. Many commodity prices are characterized by recurrent volatility switches that predate the Great Moderation period, while also exhibiting price rallies and collapses akin to business cycles. This similarity in features of the data inspires our approach to modeling commodity prices using double chain hidden Markov models.

We propose a way to model and consistently estimate parameters associated with price booms and busts in the presence of volatility switches. We apply the model to three industrial commodity prices: Crude Oil, Aluminum and Rubber, over almost 40 years of quarterly data and show that it effectively captures boom–bust related shifts in price growth rates across high–low volatility phases. The novelty of our approach is in allowing for the persistence of price changes to vary across the volatility regimes. While the lag–structure selection approach is not new, there is no model in the related economic or statistical literature that leads to independent volatility shifts originating the autoregressive structure of the underlying time series which is driven by another hidden Markov process. We let the AR structure depend on the variance regimes since, as we show in Appendix

23These prices represent the three major commodity market sectors: energy, metals and agriculture.
3. DOUBLE MIXTURE

B, the average growth rates over different regimes are independent of the lag-structure but the variances are not.

After specifying this model, we describe a likelihood profiling and filtering technique that exploits the natural Bayesian structure of hidden Markov models to “integrate out” the “incidental” volatility parameters and facilitate a two-stage estimation procedure originally due to Sir Ronald Fisher (see e.g. Sartori, 2003; Hald, 2006, Sec. 19.3). In the first stage, we estimate the parameters of primary interest (the mean price growth rates across booms and busts). In the second stage, we hold fixed the first stage parameters and estimate the volatility related incidentals. This approach obtains unbiased consistent estimates of variances while identifying the independent but recurrent volatility switches and their associated price growth persistence coefficients. Finally, in the implementation of this mixing lag–structure model, we account for the possible change in the duration and size of booms/busts by allowing, respectively, the transition probabilities and growth rates, to vary across the volatility regimes.

The remainder of the paper is organized as follows. In Section 3.2, we revisit the standard Markov switching model and provide an overview of the different approaches that have been developed to deal with persistence and time varying volatility in subsection 3.2.1. We then discuss the bias in unconditional variance estimates from these models when using maximum likelihood estimation in subsection 3.2.2. In Section 3.3, we specify a new type of model, “A Double Mixture Autoregressive Model” and describe a likelihood profiling method that affords estimation and filtering. In Section 3.4, we give results of estimating the model for three industrial commodity price time series. The final section concludes with a discussion of the results and applications in other areas of econometrics.

3.2 Review of Markov Switching Models

Since the publication of Hamilton’s (1989) influential paper on the analysis of nonstationary time series, the two state Markov switching model (MS) has found multiple applications in different areas of macroeconomics including recession forecasting (Barnett, Chauvet and Leiva-Leon, 2016), the modeling of monetary policy shifts (Sims and Zha, 2006), interest rates, stock prices, unemployment rates (Hamilton, 2016) and the identification of structural vector auto-regressions via heteroskedasticity (Lütkepohl and Velinov, 2016). MS models are also known as hidden Markov models (HMMs) and used for example in speech recognition and DNA analysis (Rabiner, 1989; Bartolucci and Pandolfi, 2013). The interest in MS/HMMs is based on their ability to approximate a general class of density functions with a wider range of values for the skewness and kurtosis than would be possible when using a single distribution. For many economic applications, these models are generally specified as a mixture of normal densities. Timmermann (2000) has shown that the mixing property enables them to generate a wide range of coefficients of skewness, kurtosis and serial correlation while using a small number of states. Recessions and

24More recently, the length and volatility of commodity bust cycles have been increasing (Roberts, 2009).
3.2. Review of Markov Switching Models

booms, policy changes or some suitably defined changes in economic time series can be parsimoniously represented by HMMs by letting the mean, variance, and possibly the dynamics of the series depend on the realization of a finite number of discrete states.

The basic MS model discussed by Hamilton (1989) models the growth of rate of a time series \( y_t \) as the outcome of a first order hidden Markov chain with states \( S_t = \{0, 1\} \) defined by transition probabilities:

\[
p_{j|i} = \text{Prob} \left[ S_t = j | S_{t-1} = i \right], \quad i, j = 0, 1
\]

The time series \( y_t \), follows the auto-regression:

\[
y_t = \mu_{S_t} + \sum_{l=1}^{\ell} \phi_l (y_{t-l} - \mu_{S_{t-l}}) + e_t, \quad e_t \sim \text{i.i.d } N(0, \sigma^2_e) \tag{3.2}
\]

where mean growth rate \( \mu_{S_t} \) depends on current state \( S_t \) and realizations of the state going back to a maximum of \( \ell \) lags. The presence of lags in the dynamic regression can be dealt with by defining a new hidden state variable together with a transition matrix. For example, if \( \ell = 1 \) so that we have only one lag, we can define a new four state Markov chain with states \( S^*_t \) and transition matrix \( P^* \):

\[
S^*_t = \begin{cases} 
0 & \text{if } S_t = 0, S_{t-1} = 0 \\
1 & \text{if } S_t = 1, S_{t-1} = 0 \\
2 & \text{if } S_t = 0, S_{t-1} = 1 \\
3 & \text{if } S_t = 1, S_{t-1} = 1 
\end{cases}
\]

\[
P^* := \begin{pmatrix} 
P_{0|0} & P_{1|0} & 0 & 0 \\
0 & 0 & P_{0|1} & P_{1|1} \\
P_{0|0} & P_{1|0} & 0 & 0 \\
0 & 0 & P_{0|1} & P_{1|1} 
\end{pmatrix}
\tag{3.3}
\]

The state variable \( S^*_t \) now keeps track of the previous period’s state but is still the outcome a first-order Markov chain whose transition matrix \( P^* \) preserves the normalization \( \sum_{j=0}^{3} P_{j|i} = 1 \) for \( i = \{0, 1, 2, 3\} \). After expanding the state-space and transition matrix to account for the presence of lags, estimation requires a specification of the likelihood function which can be efficiently evaluated using the filtering procedures of Hamilton (1990); Kim and Nelson (1999).

### 3.2.1 State Dependent Volatility and Persistence

In equation (3.2), the lag structure \( \ell \) is independent of the state and the innovation variance is a constant \( \sigma^2_e \). In the econometric literature, the most common change to this specification usually involves letting the innovation variance be state dependent: \( \sigma^2_e = \sigma^2_{S_t} \) (e.g. Hamilton and Lin, 1996). A popular extension when the data under consideration are financial time series such as stock prices or interest rates involves the introduction of ARCH or GARCH type dynamics to the volatility (Cai, 1994; Fong and See, 2001; Marcucci, 2005). In some cases, the AR coefficients are also allowed to vary across the states: in equation (3.2), \( \phi_l \) becomes \( \phi_{S_t,l} \). Psaradakis and Spagnolo (2006) study a Markov mixture of \( k \) linear autoregressive models, each of order \( \ell \), with a smooth transition type of switching
between the different states. More recently, Chang, Choi and Park (2017) develop a new type of regime switching model where the regimes are determined by an autoregressive latent factor exceeding some threshold. The latent factor is allowed to be correlated to the innovation of the observed time series, so this model can be interpreted as a form of volatility modeling. These extensions using economic or financial time series rarely let the lag-length vary across regimes.

In the business cycle dating literature, it has been known that the standard MS model breaks down when the data are extended to periods after 1984; it fails to identify the transitions from booms to recessions that characterize business cycles across most of the world’s industrialized economies (Boldin, 1996; Chauvet and Su, 2014). Business cycle forecasters have identified the need to model the fall in volatility of many macroeconomic time series, now generally referred to as the “Great Moderation”\(^{25}\), in order to maintain the model’s empirical relevance when using time series after 1984. The most interesting approach and the focus of this paper, has been the introduction of a second Markov chain to model the downward shift in volatility during Great Moderation era. An early application of this approach to modeling economic growth with two independent Markov chains that captured the Great Moderation was implemented by McConnell and Perez Quiros (2000). In their textbook implementation of Hamilton’s (1989) model, Kim and Nelson (1999) added a dummy variable for growth rates after 1984 to potentially capture “a change in the mean growth rates during boom or recession” (p. 80). Lettau, Ludvigson and Wachter (2008) use independent chains to model the lower volatility of consumption growth while examining how lower risk led to the stock market boom of the 1990s. More recent implementations of this approach include Chen (2006); Bai and Wang (2011); Doornik (2013) and Chauvet and Su (2014) who also provide a comprehensive survey.

In general, implementations of the MS model for macroeconomic data that correctly identify recession dates in the modern era require two independent Markov chains: one for the switching mean growth rates and a second chain for the variance with an “absorbing” low volatility state that occurs around 1984. In all these specifications, the lag length reduces to one (Chen, 2006; Doornik, 2013) or zero (Bai and Wang, 2011). The failure to properly account for the changing persistence of the data across the volatility phases has large consequences for the precision of implied unconditional variances.

### 3.2.2 Finite Sample Biases

Using asymptotic theory, it has been shown that maximum likelihood estimators (MLEs) of MS models are consistent and unbiased (Krishnamurthy and Ryden, 1998; Douc, Moulines and Rydén, 2004; Franke, 2012). However, simulation studies suggest that MLEs of these models as actually characterized by large biases (see e.g. Psaradakis and Sola, 1998; Ho, 2001). Psaradakis and Sola (1998) investigated the finite-sample properties of ML estimators in a first-order autoregressive model subject to Markov shifts in the mean and con-

\(^{25}\)See e.g. Justiniano and Primiceri (2008) and the references therein.
3.3. A Double Mixture Autoregressive Model

They show that obtaining estimates close enough to the true data generating process requires samples longer or larger than what is typically available for macroeconomic time series – a minimum of 400 observations should be available in empirical applications. Similarly, Ho (2001) shows that a bootstrap MLE is required in order to ameliorate the small sample biases that arise from switching regressions with Markov processes.

This bias in MS models can be linked to an old statistical result due to Neyman and Scott (1948) on the inconsistency of MLEs when dealing with partially consistent observations/data. In these models, one is dealing with observation sequences that obey a probability law with different parameter values, similar to Example 2 in Lancaster (2000) or Basu (2011) and Example 3 in Neyman and Scott (1948). In these illustrations, the authors show the inconsistency of maximum likelihood estimators of the variance or “structural parameter” when the data follow $k$ probability laws. In the context of MS models, we are trying to estimate a common $\sigma$ from $k$ distributions with $\mu_i \neq \mu_j \forall i, j = 1, \ldots, k$. If only variances are switching as in Lanne, Lütkepohl and Maciejowska (2010), then we want to estimate a common $\mu$ from $k$ distributions with $\sigma_i \neq \sigma_j \forall i, j = 1, \ldots, k$. In this case, it can again be shown that MLEs are inconsistent. This leads us to specify a new model and estimation procedure that overcome these problems.

### 3.3 A Double Mixture Autoregressive Model

Let $y_t = \Delta \log(P_t), t = 1 : T$ represent a time series of the change in the log price of a commodity. Let $S^m_t$ and $S^\nu_t$ represent, respectively, the mean and variance regime indicators. Here $S^\nu_t = \{0, 1\}$ captures volatility changes characterizing many commodity price series while $S^m_t = \{0, 1\}$ represents shifts in the growth rate related to price boom–bust cycles. The regimes $S^m_t$ and $S^\nu_t$ are each the outcome of an independent first order Markov chain with transition matrices: $P^m = p^m_{ij}$ and $P^\nu = p^\nu_{ij}$, respectively. The two components correspond to a restricted four regime model, with state $S_t = S^m_t \times S^\nu_t$ and transition matrix:

$$
\begin{array}{c|c|c|c}
S^\nu_t & S^m_t = 0 & S^m_t = 1 \\
\hline
S^m_t = 0 & S_t = 0 & S_t = 1 \\
S^m_t = 1 & S_t = 2 & S_t = 3 \\
\end{array}
$$

(3.4)

The state $S_t$ defines a dynamic linear model:

$$
y_t = \begin{cases} 
\mu_{S^m_t} + \sum_{l=1}^{\ell(S^m_t)} \phi_l \left( y_{t-l} - \mu_{S^m_{t-l}} \right) + \sigma_{S^m_t} e_t, & S^m_t = \{0, 1\}, \\
\mu_{S^\nu_t} + \sum_{l=1}^{\ell(S^\nu_t)} \phi_l \left( y_{t-l} - \mu_{S^\nu_{t-l}} \right) + \sigma_{S^\nu_t} e_t, & S^\nu_t = \{0, 1, \ldots, K\} \\
\end{cases}
\quad e_t \sim \text{iid } N(0, 1)
$$

(3.5)
with time varying intercepts $\mu_{S_t}$ and volatilities $\sigma_{S_t}$. The lag length $\ell(S_t^0)$ in (3.5) is potentially changing across variance regimes. This is the first innovation in the paper. In the original model of Hamilton (1989), there are no volatility changes with the state and $\sigma$ can be thought of as a nuisance parameter, in the sense of Sartori (2003) or Elliott, Müller and Watson (2015), which we are not interested in. In the present context, we are interested in modeling the boom–bust related shifts in mean growth rates while treating the change in volatility as incidental shift parameters in the sense of Neyman and Scott (1948, Example 1).\footnote{Note that we could alternatively treat the $\mu$s as the nuisance/incidental parameters if we were interested in estimating the $\sigma$s}

In the next two subsections, we describe an estimation and filtering procedure that implements Basu’s suggestion “to fix a prior, compute the posterior, integrate out the nuisance parameter from the posterior, to arrive at the posterior marginal distribution of the parameter of interest, and then let the statistical argument rest on the posterior marginal distribution” (Basu, 2011, paragraph 10, Section 1). We describe the two–stage MLE procedure in the third subsection.

### 3.3.1 Likelihood function

In the standard ML approach, if $S_t$ were observed, the parameters $[\mu_{S_t}, \sigma_{S_t}]_{S_t=0}^3$ of equation (3.5) would be estimated by maximizing the log-likelihood function $\ln L = \sum_{t=1}^{T} \ln f(y_t|S_t)$ where $f(y_t|S_t)$ is the conditional distribution of $y_t$ given the state $S_t$:

$$f(y_t|S_t) = \frac{1}{\sqrt{2\pi} \sigma_{S_t}} \exp \left( -\frac{1}{2} \frac{(y_t - \mu_{S_t} - \sum_{l=1}^{\ell} \psi_l (y_{t-l} - \mu_{S_{t-l}}))^2}{\sigma_{S_t}} \right).$$

However since the state $S_t$ is not observed, one would begin by considering the joint density of $y_t$, $S_t$ and information up to time $t-1$, denoted by $\psi_{t-1} = y_{1:t-1}$. The marginal density $f(y_t|S_t, \psi_{t-1})$ used in the likelihood, is then obtained by integrating out $S_t$ from $f(y_t, S_t|\psi_{t-1})$:

$$f(y_t|\psi_{t-1}) = \sum_{S_t} f(y_t|S_t, \psi_{t-1}) f(S_t|\psi_{t-1}) = \sum_{K=0}^{K=4} f(y_t|S_t, \psi_{t-1}) P[S_t = k|\psi_{t-1}]$$

where $P[S_t = k|\psi_{t-1}]$, is the prior state probability\footnote{See Kim and Nelson (1999, p. 63) for textbook illustrations of deriving the updating formula.} and $f(y_t|\psi_{t-1})$ is the posterior likelihood having observed $y_t$. For the $K = 4$ state case described by (3.4) without lags the log-likelihood

$$\ln L = \sum_{t=1}^{T} \ln \left( \sum_{K=0}^{4} f(y_t|S_t, \psi_{t-1}) \times P[S_t = k|\psi_{t-1}] \right)$$

would be maximized to obtain the parameters $\theta = \{p_{0|0}^m, p_{1|1}^m, p_{0|0}^v, p_{1|1}^v, \mu_0, \mu_1, \sigma_0, \sigma_1\}$.

To implement Basu’s (2011) idea of integrating out the nuisance parameter, start with (3.6) having fixed a prior $P[S_t = k|\psi_{t-1}]$ and compute the posterior $f(y_t|S_t, \psi_{t-1})$. We now want to integrate out the incidental parameter $\sigma_{S_t}$ to arrive at the posterior marginal distribution.
3.3. A Double Mixture Autoregressive Model

distribution of the parameter of interest \( \mu_{S^m_t} \):

\[
f(y_t | S^m_t, \psi_{t-1}) = \sum_{k=0}^{1} f(y_t | S^m_t, S^u_t = k, \psi_{t-1}) \times P|S^u_t = k|S^m_t, \psi_{t-1}
\]  

(3.7)

where the predictive density for the mean is:

\[
f(y_t | S^m_t, S^u_t = k, \psi_{t-1}) = \sum_{k_m} f(y_t | S^m_t = k_m, S^u_t = k, \psi_{t-1}) \times P|S^u_t = k_m|S^m_t, \psi_{t-1}
\]  

(3.8)

with \( k_m \) indexing the (possibly expanded) number of states of \( S^m_t \) in the presence of lags (see Section 3.3.2 below). The final likelihood is given by:

\[
\ln L = \sum_{t=1}^{T} \ln \left( \sum_{k=0}^{1} f(y_t | S^m_t, S^u_t = k, \psi_{t-1}) \times P|S^u_t = k|S^m_t, \psi_{t-1} \right).
\]  

(3.9)

3.3.2 Filtering

In order to evaluate equations (3.7) and (3.8), we need to compute the prior probabilities \( P|S^m_t = k_m|S^u_t = k, \psi_{t-1} \) and \( P|S^u_t = k|S^m_t, \psi_{t-1} \). Given a value of the prior \( P|S_t = k|\psi_{t-1} \) and the likelihood \( f(y_t | S_t = l, \psi_{t-1}) \), we can compute the posterior, \( P(S_t = k|\psi_t) \) using Bayes’s law as:

\[
P(S_t = k|\psi_t) = \frac{P(S_t = k|\psi_{t-1})f(y_t | S_t = k, \psi_{t-1})}{f(y_t | \psi_{t-1})}
\]  

(3.10)

where \( \psi_t = \{\psi_{t-1}, y_t\} \). Expanding the state space to account for the number of lags gives \( n_m = 2^{(|S^u_t|+1)} \) states for each \( m = S^u_t = \{0, 1\} \) volatility state. Collect the set of possible conditional densities \( f(y_t | S^m_t, S^u_t = k) \) in a \( (1 \times n_m) \) row vector \( \eta_{m,t} \). Likewise, define the \( (1 \times n_m) \) vector \( \xi_{m,t|t-1} \) whose \( n_{m,t}^{th} \) element is \( P|S^m_t = k_m|S^u_t = k, \psi_{t-1} \). Taking the inference on \( \xi_{m,t|t-1} \), the update \( \xi_{m,t|t} \) can be calculated after observing \( y_t \). This is accomplished by using the transition matrix \( P^m_{m} \) and applying (3.10) to obtain the predicted and filtered probabilities:

\[
\xi_{m,t|t-1} = \xi_{m,t|t-1} \odot P^m_{m} \]  

(3.11)

\[
\xi_{m,t|t} = \frac{\xi_{m,t|t-1} \odot \eta_{m,t}}{\sum_{i=1}^{n_m} \xi_{m,t|t-1} \odot \eta_{m,t}^{i}} \]  

(3.12)

where \( P^m_{m} \) is an expansion similar to (3.3) of the state matrix \( P^m_{m} \) defined by (3.4), \( \odot \) is the Hadamard product and the summation is over \( n_m \) states. To update \( P|S^u_t = k|S^m_t, \psi_{t-1} \), collect these probabilities into the \( (4 \times 1) \) vector \( \xi_{u,t|t-1} \) and collapse each \( \eta_{m,t} \) over their respective lags to obtain the likelihood \( f(y_t | S^m_t, \psi_{t-1}) \) which are collected in a \( (4 \times 1) \) vector \( \eta_{u,t} \). Again, using the transition matrix and applying (3.10), we obtain the predicted and
The elements of the transition matrices $P$ are specified by:

$$
p^{m}_{0,0} = \frac{\exp(p_0)}{1 + \exp(p_0)}, \quad p^{m}_{1,0} = \frac{\exp(p_1)}{1 + \exp(p_1)}, \quad \text{for } P^m_0
$$

$$
p^{m}_{1,1} = \frac{\exp(p_{1d})}{1 + \exp(p_{1d})}, \quad p^{m}_{1,1} = \frac{\exp(p_{1d})}{1 + \exp(p_{1d})}, \quad \text{for } P^m_1
$$

$$
p^{v}_{0,0} = \frac{\exp(q_0)}{1 + \exp(q_0)}, \quad p^{v}_{1,1} = \frac{\exp(q_1)}{1 + \exp(q_1)}, \quad \text{for } P^v
$$

where $p_0, p_1, p_{0d}, p_{1d}, q_0, q_1$ are unconstrained parameters and the transition probabilities are constrained within the $[0, 1]$ interval. The two transition matrices $P^m_1$ and $P^m_2$ in (3.15) allow for the possibility of different transition probabilities across the volatility states $S_t^m$; with the mean growth rates in the second volatility state being: $\mu_{0d}, \mu_{1d}$.

The likelihood (3.9) is maximized over the parameter vector $\theta = \{p_0, p_1, p_{0d}, p_{1d}, q_0, q_1, \mu_0, \mu_1, \mu_{0d}, \mu_{1d}, \sigma_0, \sigma_1, \phi_0^0, \phi_0^1, \phi_1^0, \phi_1^1, \phi_{\ell_0}^0, \phi_{\ell_1}^1\}$ where $\ell_0$ need not equal $\ell_1$, in two stages. In the first stage, suppress the volatility related autoregressive parameters while estimating the boom–bust and volatility switch parameters: $\theta = \{p_0, p_1, p_{0d}, p_{1d}, q_0, q_1, \mu_0, \mu_1, \mu_{0d}, \mu_{1d}, \sigma_0, \sigma_1\}$ which gives consistent estimates of the parameters: $\theta' = \{p_0, p_1, p_{0d}, p_{1d}, q_0, q_1, \mu_0, \mu_1, \mu_{0d}, \mu_{1d}\}$.

In the second stage, hold the consistently estimated parameters $\theta'$ fixed and use the profile likelihood to estimate second stage parameters $\ell = \{\sigma_0, \sigma_1, \phi_0^0, \phi_0^1, \phi_1^0, \phi_1^1\}$.

### 3.4 Empirical Results

The time series analyzed come from the IMF database of commodity prices, available at imf.org. The time series are of monthly frequency from Jan. 1980 to May 2019. We con-

---

vert the series to quarterly frequency by taking the last value of each quarter. Figures 3.1a, 3.1b and 3.1c shows the time variation of the log prices we analyze. The shaded regions are United States NBER recession dates. The solid lines represent periods when the change in the price series exceeds the average over the sample period plus one standard deviation. These high volatility periods tend to occur around recessions and are particularly prominent around the Great Recession of 2007–2008.

**Figure 3.1: log Prices**

(a) Crude Oil

![Crude Oil](image)

(b) Aluminum

![Aluminum](image)

(c) Rubber

![Rubber](image)

The dashed blue line represents log prices for each commodity. The shaded regions are United States NBER recession dates. The highlighted solid parts of the lines represent periods when the change in price $\Delta y_t$ exceeds the average change in price plus one standard deviation. The respective volatilities for each line segment are given by $\sigma(\Delta y_t)$

3.4.1 First Stage Estimates: $\mu_m$

We estimate the model described in Section 3.3 using the non-linear programming solver `fminunc` in MATLAB R2017b. Standard errors are obtained from the inverted Hessian matrix using the delta method. Table 3.1 gives results of the first stage estimation. To aid in
interpretation of these results, the parameter estimates should be interpreted in line with illustrations of state inferences in Figures 3.2–3.4. Column (1) of Table 3.1 gives estimates of parameters that describe the boom–bust cycles and volatility switches that characterize the price of Crude Oil. $\mu_0 < 0$ means that the state $S^m_t = 0$ captures a period of falling prices. This can be seen in Figure 3.2 which shows the downward trend in prices from the beginning of the sample till around 1999. This is followed by a persistent upward trend that has been ongoing over the last 20 years, with a brief interruption occurring around the recession years 2007–2008. The next parameter estimate $\mu_1 > |\mu_0|$ implies that price booms are on average larger than busts. The boom and bust periods divide the whole sample period into two almost equal parts ($p_{0,0}^m \approx p_{0,1}^m$), but since $\mu_1 > |\mu_0|$ we end up with a price higher than that at the beginning of the sample. There are recurrent periods of high volatility, especially around recessions. During such periods, the oil price volatility increases fourfold $\sigma_1 \approx 4\sigma_0$. This phenomenon is observed the other two commodity prices as well.

Column (2) in Table 3.1 shows the estimates for Aluminum prices. In this case the estimate $\mu_0 > 0$ implies the state $S^m_t = 0$ is a price boom. This can be seen in the upper part of Figure 3.3, which plots the state inferences $\text{Prob}[S^m_t]$ vis-à-vis the price series. The states are well defined, with frequent transitions between booms and busts. The booms are relatively short, lasting approximately $1 = \frac{1}{1 - p_{0,0}^m} = 6$ quarters during periods of low volatility and just around 4 months under the high volatility state. Since the brief recession of July 1990 to April 1991, there is only one episode of high volatility prices which occurs just after the onset of the Great Recession in 2007. Finally, Column (3) shows the estimates from Rubber prices. These are similar in behavior to those of Aluminum in their boom–bust cycles: there are multiple transitions between price rises and falls, with the boom periods generally shorter than the busts. However, there are more frequent high volatility periods which are independent of United States recessions. This may be an artifact of Rubber being an agricultural–industrial commodity whose production is also affected by weather related events.

### 3.4.2 Second Stage Estimates: $\sigma_\upsilon$

We now hold fixed the set of parameters relating to the $\mu_m$s and re-estimate the model. The results in Table 3.2 are the best model estimates based on the AutoMetrics approach of starting with the most general model where $(\ell_0, \ell_1) = (4,4)$ and testing for significance of individual coefficients (Doornik and Hendry, 2014) with the final model selected using the lowest values of the standard information criteria. This process results into models of different lag-structures for each time series: (4,2) for Crude Oil, (4,4) for Aluminum and (4,3) for Rubber. The last six rows of Table 3.2 show the results of AutoMetrics style testing to get the final model shown. In the case of Aluminum in column (2), all AR coefficients are significant up to lag-4 across the volatility regimes. For Crude Oil and Rubber, the likelihood ratio tests against alternatives with more AR parameters fail to reject the null as
3.4. Empirical Results

Table 3.1: First Stage Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate (1)</th>
<th>S.E. (1)</th>
<th>Estimate (2)</th>
<th>S.E. (2)</th>
<th>Estimate (3)</th>
<th>S.E. (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_0 )</td>
<td>-0.0146</td>
<td>0.0014</td>
<td>0.0669</td>
<td>0.0026</td>
<td>0.0470</td>
<td>0.0016</td>
</tr>
<tr>
<td>( \mu_1 )</td>
<td>0.0410</td>
<td>0.0011</td>
<td>-0.0421</td>
<td>0.0012</td>
<td>-0.0335</td>
<td>0.0007</td>
</tr>
<tr>
<td>( \mu_{0d} )</td>
<td>-0.0013</td>
<td>0.0727</td>
<td>0.0553</td>
<td>0.0078</td>
<td>-0.0323</td>
<td>0.0027</td>
</tr>
<tr>
<td>( \mu_{1d} )</td>
<td>-0.1046</td>
<td>0.0050</td>
<td>-0.2160</td>
<td>0.0077</td>
<td>-0.0010</td>
<td>0.0319</td>
</tr>
<tr>
<td>( \sigma_0 )</td>
<td>0.0949</td>
<td>0.0009</td>
<td>0.0568</td>
<td>0.0008</td>
<td>0.0565</td>
<td>0.0006</td>
</tr>
<tr>
<td>( \sigma_1 )</td>
<td>0.3185</td>
<td>0.0038</td>
<td>0.1403</td>
<td>0.0030</td>
<td>0.1974</td>
<td>0.0017</td>
</tr>
<tr>
<td>( P_{000}^m )</td>
<td>0.9788</td>
<td>0.0021</td>
<td>0.6484</td>
<td>0.0108</td>
<td>0.9626</td>
<td>0.0055</td>
</tr>
<tr>
<td>( P_{011}^m )</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.8263</td>
<td>0.0059</td>
<td>0.9803</td>
<td>0.0021</td>
</tr>
<tr>
<td>( P_{100}^m )</td>
<td>0.6189</td>
<td>0.4379</td>
<td>0.8064</td>
<td>0.0195</td>
<td>0.9998</td>
<td>0.0430</td>
</tr>
<tr>
<td>( P_{111}^m )</td>
<td>1.0000</td>
<td>0.8625</td>
<td>0.4684</td>
<td>0.0203</td>
<td>0.6802</td>
<td>0.0297</td>
</tr>
<tr>
<td>( P_{000}^v )</td>
<td>0.8936</td>
<td>0.0040</td>
<td>0.9362</td>
<td>0.0040</td>
<td>0.8098</td>
<td>0.0062</td>
</tr>
<tr>
<td>( P_{011}^v )</td>
<td>0.6289</td>
<td>0.0113</td>
<td>0.6505</td>
<td>0.0194</td>
<td>0.7951</td>
<td>0.0083</td>
</tr>
</tbody>
</table>

| Dur.: \( S_t^m = 1 | S_t^v = 0 \) | 131.2895 | 5.7560 | 50.8877 |
| Dur.: \( S_t^m = 1 | S_t^v = 1 \) | 25.7105  | 1.1853 | 1.0426  |
| Log Likelihood  | 76.5943    | 142.3014 | 100.4833 |
| AIC              | -0.8229    | -1.6599  | -1.1272  |
| BIC              | -0.0202    | -1.0398  | -0.5071  |
| HIC              | -0.7280    | -1.5650  | -1.0323  |
| T                | 157        | 157      | 157      |

Comparing the second stage estimation results in Table 3.2 to the first stage estimates in Table 3.1, we see that in columns (1) & (3), \( \sigma_0, \sigma_1 \) and \( P_{001}^m, P_{111}^m \) remain about the same as before for Crude Oil and Rubber, but with significant autoregressive coefficients across the volatility regimes. For Aluminum prices, the estimates in Column (2) significantly differ from their first stage counterparts. The volatility estimates change from \( \sigma_0 = 0.0568, \sigma_1 = 0.1403 \) in the first stage to \( \sigma_0 = 0.0391, \sigma_1 = 0.1120 \) in the second stage, while the transition probabilities change from \( P_{001}^v = 0.94, P_{111}^v = 0.65 \) to \( P_{001}^v = 0.85, P_{111}^v = 0.45 \).

3.4.3 State Inference Using 2nd Stage Estimates

Figures 3.5a–3.5c gives the smoothed state inference of the boom-bust states \( S_t^m \) given second stage estimates. As can be seen, using the second stage estimates, the boom/bust state inferences do not qualitatively change, but the probabilities are more well defined: dotted second stage lines are closer to 0/1 than the solid first stage.

Figures 3.6a–3.6c, show inference on the state \( S_t^v = 1 \) after the second stage where the volatility related parameters and AR coefficients have been re-estimated. The second stage inference in qualitatively identical to the first stage in Figures 3.2–3.4, which is shown in the rows labeled L.R. Test.
what we would expect given that the only inconsistent parameter from the first–stage are
the $\sigma_\nu$ – the transition probabilities $P_{01}^\nu$ and $P_{10}^\nu$ are not affected by the incidental parame-
ter problem. The dotted lines represent the volatility state inference after second stage,
$\Pr(S_\nu^t = 1, 2)$, while the solid lines represent the same from first stage. While qualita-
tively the same, second stage inference is more well defined: dotted lines are closer to the
0/1 edges than solid first stage lines.

As before, the log-price changes are plotted with the probabilities to show how well
the model identifies periods of high volatility. Using the simple rule of assigning an obser-
vation to a state when the probability exceed one half, there are some switches between
first and second stage estimation. In Figures 3.6a–3.6c, these are shown as the points with
upward–blue and downward–black triangles, representing respectively, addition and re-
moval from the high volatility state in comparison to first stage assignment. This results
into the identification of more price spike instances for the Aluminum and Rubber prices.
The average duration/number of high/low volatility period do not change in the case of
Aluminium as the additions and removals are approximately the same in number: 9 versus
7 – in line with the similarity of transition probabilities from first to second stage estima-
tion. For Rubber, there are 18 additions to the high volatility state and 10 removals – but
in this case, the transition probabilities do not change much, so the result could be due to
accounting for the persistence of the boom–bust cycles.

Finally, Table 3.3 shows the unconditional moments summarized over the volatility
regimes. We use the unconditional probabilities in the rows under $S_t^\nu$ and the respective
3.4. Empirical Results

**Figure 3.3:** First Stage Smoothed States Inference: Aluminum

Notes: See notes below Figure 3.2.

μₘₜ to compute: \( \frac{\mu_0}{\mu_0 + \mu_1} \) when \( S^o_t = 0 \) and \( \frac{\mu_0}{\mu_0 + \mu_1} \) when \( S^o_t = 1 \). We then compute the unconditional means and variances using equations (B.2) and equation (B.11) shown in Appendix B. Our model therefore collapses to a form with a single state variable \( S^o_t \), where \( y_t = \mu_{S^o_t} + \sum_{l=1}^{f(S^o_t)} \Phi_l S^o_t \left( y_{t-l} - \mu_{S^o_{t-l}} \right) + \sigma_{S^o_t} \epsilon_t \), \( S^o_t \in \{0, 1\} \). This form of MS model is frequently found in applications involving monetary policy, interest rate and for identification by heteroskedasticity (Netsunajev, 2013; Lütkepohl and Velinov, 2016), except for the time varying number of lags.

The rows under \( S^o_t \) in Table 3.3 therefore show the moments when we have collapsed the model to a single chain with two states and switching in both mean and variance. For Crude Oil, high volatility is accompanied on average by price falls with every recession since 1990 is accompanied by an instance of sharp price falls. The reverse is true for Aluminum: the high volatility state is on average during period when the price is rising quickly, while the low volatility state occurs during period of modest, almost zero, average price falls. Finally, Rubber prices exhibit high volatility when prices are on average falling.

The unconditional probabilities of being in a high/low volatility state are given by the last rows of Table 3.3. These results show the heterogeneity in the behavior of commodity prices across volatility regimes that has possibly not been documented before.
Figure 3.4: First Stage Smoothed States Inference: Rubber

Notes: See notes below Figure 3.2.
### Table 3.2: Second Stage Estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>(1) AR(4,2)</th>
<th>(2) AR(4,4)</th>
<th>(3) AR(4,3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodity</td>
<td>Crude Oil</td>
<td>Aluminum</td>
<td>Rubber</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.0943</td>
<td>0.0391</td>
<td>0.0532</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.2981</td>
<td>0.1120</td>
<td>0.1834</td>
</tr>
<tr>
<td>$p_{\mu0}^0$</td>
<td>0.8934</td>
<td>0.8446</td>
<td>0.7697</td>
</tr>
<tr>
<td>$p_{\mu11}^0$</td>
<td>0.6129</td>
<td>0.4462</td>
<td>0.7883</td>
</tr>
<tr>
<td>$\phi_1^1$</td>
<td>-0.0397</td>
<td>0.1245</td>
<td>0.4121</td>
</tr>
<tr>
<td>$\phi_2^1$</td>
<td>-0.1703</td>
<td>0.3224</td>
<td>-0.0650</td>
</tr>
<tr>
<td>$\phi_3^1$</td>
<td>-0.0157</td>
<td>-0.0422</td>
<td>-0.1002</td>
</tr>
<tr>
<td>$\phi_4^1$</td>
<td>-0.0299</td>
<td>-0.0697</td>
<td>-0.1498</td>
</tr>
<tr>
<td>$\phi_1^2$</td>
<td>-0.0506</td>
<td>0.3443</td>
<td>0.0095</td>
</tr>
<tr>
<td>$\phi_2^2$</td>
<td>-0.4591</td>
<td>-0.3029</td>
<td>-0.2102</td>
</tr>
<tr>
<td>$\phi_3^2$</td>
<td>-0.0800</td>
<td>-0.4169</td>
<td>0.1913</td>
</tr>
<tr>
<td>$\phi_4^2$</td>
<td>0.4206</td>
<td>27.0603</td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>78.2562</td>
<td>149.0757</td>
<td>99.8587</td>
</tr>
<tr>
<td>L.R. Test</td>
<td>0.2145</td>
<td>0.2145</td>
<td>0.2145</td>
</tr>
<tr>
<td>LR $p$-value</td>
<td>0.6432</td>
<td>0.6432</td>
<td>0.6432</td>
</tr>
<tr>
<td>AIC</td>
<td>-0.7877</td>
<td>-1.6873</td>
<td>-1.0570</td>
</tr>
<tr>
<td>BIC</td>
<td>0.1607</td>
<td>-0.6336</td>
<td>-0.0560</td>
</tr>
<tr>
<td>HIC</td>
<td>-0.6428</td>
<td>-1.5263</td>
<td>-0.9041</td>
</tr>
</tbody>
</table>

### Table 3.3: Unconditional Moments

<table>
<thead>
<tr>
<th>State</th>
<th>Moment</th>
<th>(1) Crude Oil</th>
<th>(2) Aluminum</th>
<th>(3) Rubber</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{t}^m = 0, 1$</td>
<td>$\mu_0$</td>
<td>-0.0146</td>
<td>0.0669</td>
<td>0.0470</td>
</tr>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>0.0410</td>
<td>-0.0421</td>
<td>-0.0335</td>
</tr>
<tr>
<td></td>
<td>$\mu_{0d}$</td>
<td>-0.0159</td>
<td>0.1222</td>
<td>0.0147</td>
</tr>
<tr>
<td></td>
<td>$\mu_{1d}$</td>
<td>-0.0636</td>
<td>-0.2581</td>
<td>-0.0345</td>
</tr>
<tr>
<td></td>
<td>$\pi_0^\mu$</td>
<td>0.0000</td>
<td>0.3307</td>
<td>0.3450</td>
</tr>
<tr>
<td></td>
<td>$\pi_1^\mu$</td>
<td>1.0000</td>
<td>0.6693</td>
<td>0.6550</td>
</tr>
<tr>
<td></td>
<td>$\pi_0^d$</td>
<td>0.0000</td>
<td>0.7330</td>
<td>0.9994</td>
</tr>
<tr>
<td></td>
<td>$\pi_1^d$</td>
<td>1.0000</td>
<td>0.2670</td>
<td>0.0006</td>
</tr>
<tr>
<td>$S_{t}^v = 0, 1$</td>
<td>$\mu$</td>
<td>0.0410</td>
<td>-0.0061</td>
<td>-0.0057</td>
</tr>
<tr>
<td></td>
<td>$\mu_d$</td>
<td>-0.0636</td>
<td>0.0207</td>
<td>0.0147</td>
</tr>
<tr>
<td></td>
<td>$\sigma_0$</td>
<td>0.0092</td>
<td>0.0041</td>
<td>0.0052</td>
</tr>
<tr>
<td></td>
<td>$\sigma_1$</td>
<td>0.1127</td>
<td>0.0597</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>$\pi_0^\nu$</td>
<td>0.7841</td>
<td>0.7809</td>
<td>0.4790</td>
</tr>
<tr>
<td></td>
<td>$\pi_1^\nu$</td>
<td>0.2159</td>
<td>0.2191</td>
<td>0.5210</td>
</tr>
</tbody>
</table>
Figure 3.5: Smoothed Second Stage Bust State Inference

Notes: Smoothed State Inference is plotted on the left axis with the dotted line representing second stage \( \text{Prob}(S^m_t = 1) \) and the solid line representing the same from first stage. The dark and light circles mark \( S^m_t = 0 \) and \( S^m_t = 1 \), respectively. The upwards triangles mark instances when point has moved from boom to bust state after second stage estimation while downwards triangles mark the opposite case.
3.4. Empirical Results

Figure 3.6: Smoothened Second Stage Volatility State Inference

Notes: Smoothened State Inference is plotted on the left axis with the dotted line representing second stage \( \text{Prob}[S^v_t = 1] \) and the solid line representing the same from first stage. The dark and light circles mark \( S^v_t = 0 \) and \( S^v_t = 1 \), respectively. The upwards triangles mark instances when point has moved from the low to high volatility state after 2nd stage estimation while downwards triangles mark the opposite case.

(a) Crude Oil

(b) Aluminum

(c) Rubber
3. DOUBLE MIXTURE

3.5 Conclusion

In this paper, we developed a new type of mixture of normals model for time series subject to both changes in growth and volatility. The innovative part of the model is in allowing the persistence of a time series to be a function of its volatility. Current models that use the same number of lags across volatility regimes are essentially imposing a restriction on the persistence across volatility regimes which, as we have shown, is not always the case. By dispensing with this restriction, we developed a new approach to the modeling of time series subject to regime change. We argued that this model specification and its two-stage estimation procedure results into a better representation of the boom–bust and high–low volatility features of commodity price time series. Our model was able to successfully identify periods of boom–bust cycles in prices which are independent of the volatility phases.

The results have potential applications in a wide range of econometric analyses including the identification by heteroskedasticity literature advanced by Netsunajev (2013) and Lütkepohl and Velinov (2016) amongst others, and the asset pricing and learning literature of Lettau, Ludvigson and Wachter (2008).

---

29 This literature requires estimation of the variance across regimes, so the results here directly apply.
30 This literature requires the estimation of the variance of say consumption growth, which measures risk and determines the expected return to the aggregate stock market.
Appendix

A Spread Estimates with Autocorrelated Trades

Under Roll's original assumptions, the transactions price follows the process: \( \Delta p_t = s \Delta q_t + u^m_t \) from which it follows that the spread can be estimated as: \( s = \sqrt{-\text{Cov}(\Delta p_t, \Delta p_{t-1})} \), where Cov is the auto-covariance of price changes. The estimate of \( s \) obtained from the auto-covariance is usually called the moment estimator because it is the sample estimate for the unknown parameter. If \( q_t \) were observed, the maximum-likelihood estimator of \( s \) would be:

\[
\hat{s}_{\text{MLE}} = \frac{\text{Cov}(\Delta p_t, \Delta q_t)}{\text{Var}(\Delta q_t)} = s.
\]

Under the assumptions 1 and 2, the price process can be written as: \( \Delta p_t = \Delta m_t + \Delta(s_t q_t) = s \Delta q_t + u_t \), where \( u_t = u^m_t + \Delta(u^p_t q_t) \) with \( u^p_t \perp q_t \) and we can compute the covariance of price changes based on the information on Table A.1, as: \( \text{Cov}(\Delta p_t, \Delta p_{t-1}) = -4 \times s^2 \times p_{21}p_{12} \) and the unbiased estimator of Roll’s spread measure as:

\[
\hat{s}_{\text{Roll}} = \sqrt{-\text{Cov}(\Delta p_t, \Delta p_{t-1})} = 2 \times \hat{s}_{\text{MLE}} \times \sqrt{p_{21}p_{12}}
\] (A.1)

The entries in the last column of Table A.1 are computed using the chain rule of conditional probabilities and first-order Markov property of \( q_t \):

\[
\text{Prob}(q_{t+1}, q_t, q_{t-1}) = \text{Prob}(q_{t+1} | q_t) \text{Prob}(q_t | q_{t-1}) \text{Prob}(q_{t-1})
\]

where \( \pi_1 = \text{Prob}(q_{t-1} = 1) = \frac{p_{12}}{p_{12} + p_{21}}, \pi_2 = 1 - \pi_1 = \text{Prob}(q_{t-1} = -1) \) are the unconditional probabilities. Covariance is computed using the quantities \( \mathbb{E}(q_t) = \pi_1 - \pi_2, \mathbb{E}(\Delta q_t) = \mathbb{E}(\Delta q_{t+1}) = 0 \)\(^{31} \) and \( \mathbb{E}(\Delta q_t + \Delta q_{t+1}) = -4(p_{21}p_{12} + p_{12}p_{21}) \). Similar corrections

\(^{31}\) Using the first four rows of Table A.1, the expectations are computed: \( \mathbb{E}(\Delta q_{t+1}) = -2p_{12}\pi_1 + 2p_{21}\pi_2 = \)

<table>
<thead>
<tr>
<th>( q_{t+1} )</th>
<th>( q_t )</th>
<th>( q_{t-1} )</th>
<th>( \Delta q_{t+1} )</th>
<th>( \Delta q_t )</th>
<th>( \text{Prob}(q_{t+1}, q_t) )</th>
<th>( \Delta q_{t+1} \Delta q_t )</th>
<th>( \text{Prob}(q_{t+1}, q_t, q_{t-1}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>0</td>
<td>0</td>
<td>( p_{11} \pi_1 )</td>
<td>0</td>
<td>( p_{11}p_{11} \pi_1 )</td>
</tr>
<tr>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>-2</td>
<td>0</td>
<td>( p_{12} \pi_1 )</td>
<td>0</td>
<td>( p_{12}p_{11} \pi_1 )</td>
</tr>
<tr>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>+2</td>
<td>-2</td>
<td>( p_{21} \pi_2 )</td>
<td>-4</td>
<td>( p_{21}p_{12} \pi_1 )</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>0</td>
<td>-2</td>
<td>( p_{22} \pi_2 )</td>
<td>0</td>
<td>( p_{22}p_{21} \pi_2 )</td>
</tr>
<tr>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>+2</td>
<td>( p_{11} \pi_1 )</td>
<td>0</td>
<td>( p_{11}p_{21} \pi_2 )</td>
</tr>
<tr>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-2</td>
<td>+2</td>
<td>( p_{12} \pi_1 )</td>
<td>-4</td>
<td>( p_{12}p_{21} \pi_2 )</td>
</tr>
<tr>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+2</td>
<td>0</td>
<td>( p_{21} \pi_2 )</td>
<td>0</td>
<td>( p_{21}p_{22} \pi_2 )</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>( p_{22} \pi_2 )</td>
<td>0</td>
<td>( p_{22}p_{22} \pi_2 )</td>
</tr>
</tbody>
</table>
Appendix A. Spread Estimates with Autocorrelated Trades

can be found in Chapter 2 of Foucault et al. (2013).

\[
2 \left( \frac{p_1 p_{12} - p_{12} p_1}{p_{12} + p_{21}} \right) = 0
\]
Appendix B. Moments of a Two-State Markov Switching Model

B Moments of a Two-State Markov Switching Model

Timmermann (2000) and Petříčková (2014) have derived moments of a general class of Markov regime-switching models. Here, we derive similar moments but in the case where there are two states and the variance doesn't change with the regime (in contrast to the standard assumption in many applications of regime switching for financial market data).

For the model in (3.2) with $\ell = 0$,

$$\mu = \mathbb{E}(y_t) = \pi_0 \mu_0 + \pi_1 \mu_1$$

where the $\pi_i$ is the unconditional probability of being in state $S_t := \{0, 1\}$ and satisfies the recursion $\pi P = \pi$ for the transition matrix $P$. The second central moment is obtained by computing:

$$\mathbb{E}(y^2_t) = \mathbb{E}(y^2_t | S_t = 0) + \mathbb{E}(y^2_t | S_t = 1)$$

$$\mathbb{V}(y_t) = \mathbb{E}(y^2_t) - \mu^2$$

$$= \sigma^2 + \pi_0 \pi_1 (\mu_1 - \mu_0)^2.$$ (B.4)

The unconditional variance is therefore increasing in the persistence of the time series $(\pi_i)$ and the difference between growth rates. For the general case where $\ell \neq 0$ in (3.2), we use a state-space representation to recast the model as vector auto-regression of order one (VAR(1)):

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-\ell} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \ldots & \phi_\ell & 0 \\ 0 & \phi_1 & \ldots & \phi_\ell & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \phi_1 & \phi_2 \\ \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-\ell-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \mu S_t + \begin{bmatrix} e_t \end{bmatrix}$$

or in matrix form,

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 M_t + Ce_t.$$ (B.5)

The moments are then defined as follows:

$$M = \mathbb{E}(M_t) = 1_{\ell+1} \mu$$ (B.6)

where $1_{\ell+1}$ is a vector of ones and $\mu$ is defined by (B.2). Now define the variance-covariance
Appendix B. Moments of a Two-State Markov Switching Model

matrix of the time varying mean vector $\mathbf{M}_t$ by:

\[
\Sigma_M = \mathbb{E}(\mathbf{M}_t - \mathbf{M})(\mathbf{M}_t - \mathbf{M})'
\]

where the main diagonal entries are given by $\mathbb{E}(\mu_{S_t} - \mu)^2 = \vec{\pi}'((\vec{\mu}_s - \vec{\mu}))$ with the vectors $\vec{\pi} = (\pi_0, \pi_1)'$, $\vec{\mu}_s = (\mu_0, \mu_1)'$, $\vec{\mu} = (\mu, \mu)'$ and $\odot$ is the element by element multiplication operator. The off diagonal entries are given by $\mathbb{E}(\mu_{S_t} - \mu)(\mu_{S_{t-1}} - \mu) = \vec{\pi}'(\{P^n(\vec{\mu}_s - \vec{\mu})\} \odot (\vec{\mu}_s - \vec{\mu}))$ where the $2 \times 2$ matrix $P$ is the matrix of transition probabilities.\(^\text{32}\) Taking expectations over (B.5), we have:

\[
\mathbb{E}(\mathbf{Y}_t) = \Phi_1 \mathbb{E}(\mathbf{Y}_{t-1}) + \Phi_2 \mathbb{E}(\mathbf{M}_t)
\]

where we have used the zero mean property of $e_t$. The covariance-stationary property of $y_t$ then implies that

\[
\mathbb{E}(\mathbf{Y}_t) = (I_{\ell+1 \times \ell+1} - \Phi_1)^{-1}\Phi_2 I_{\ell+1}\mu = 1_{\ell+1}\mu
\]

To determine the unconditional variance, first subtract (B.8) from (B.5) to obtain

\[
\mathbf{Y}_t - \mathbb{E}\mathbf{Y}_t = \Phi_1(\mathbf{Y}_{t-1} - \mathbb{E}\mathbf{Y}_{t-1}) + \Phi_2(\mathbf{M}_t - \mathbf{M}) + C e_t
\]

The orthogonality of $e_t$ to $(\mathbf{Y}_{t-1} - \mathbb{E}\mathbf{Y}_{t-1})$ and $(\mathbf{M}_t - \mathbf{M})$ then implies:

\[
\mathbb{E}(\mathbf{Y}_t - \mathbb{E}\mathbf{Y}_t)(\mathbf{Y}_t - \mathbb{E}\mathbf{Y}_t)' = \Phi_1\mathbb{E}(\mathbf{Y}_{t-1} - \mathbb{E}\mathbf{Y}_{t-1})(\mathbf{Y}_{t-1} - \mathbb{E}\mathbf{Y}_{t-1})'\Phi_1'
\]

\[+ \Phi_2\mathbb{E}(\mathbf{M}_t - \mathbf{M})(\mathbf{M}_t - \mathbf{M})'\Phi_2' + C'C\sigma_e^2\]

which can be written compactly as:

\[
\Sigma_t = \Phi_1 \Sigma_{t-1} \Phi_1' + \Phi_2 \Sigma_M \Phi_2' + C'C\sigma_e^2
\]

where $\Sigma_M$ is defined by (B.7). Again, stationary $y_t$ implies $\Sigma_t = \Sigma_{t-1} = \Sigma$ so the unconditional variance is defined by:

\[
\text{vec}(\Sigma) = (I_{(\ell+1)^2} - \Phi_1 \otimes \Phi_1)^{-1}\text{vec}(\Phi_2 \Sigma_M \Phi_2' + C'C\sigma_e^2)
\]

\[= (I_{(\ell+1)^2} - \Phi_1 \otimes \Phi_1)^{-1}\left[(\Phi_2 \otimes \Phi_2)\text{vec}(\Sigma_M) + \text{vec}(C'C)\sigma_e^2\right]\]

Applying this result to the AR(1) case we would have: $\Phi_1 = [\phi_1, 0; 1, 0]$ and $\Phi_2 = [1, -\phi_1; 0, 0]$;

\(^\text{32}\)This assumes that the process generated by $P$ is time reversible, which holds in general for $2 \times 2$ chains. See Footnote 5 of Timmermann (2000) for discussion of this issue and a definition of the “backward” probabilities required in higher order chains.
Appendix B. Moments of a Two-State Markov Switching Model

$C = (1, 0)'. After some tedious algebra, it can be shown that the variance-covariance matrix $\Sigma_M$ has entries in the main diagonal $E(\mu_{S_t} - \mu)^2 = \pi_0 \pi_1 (\mu_0 - \mu_1)^2$ and off-diagonal elements $E(\mu_{S_t} - \mu)(\mu_{S_{t-1}} - \mu) = \pi_0 \pi_1 (\mu_0 - \mu_1)^2 (p_{0|0} + p_{1|1} - 1)$. This gives the variance expression

$$\Sigma_{11} = \text{Var}(y_t) = \frac{\pi_0 \pi_1 (\mu_0 - \mu_1)^2}{1 - \phi_1^2} \left[ 1 - \phi_1^2 - 2\phi_1 (p_{0|0} + p_{1|1} - 1) \right] + \frac{\sigma_e^2}{1 - \phi_1^2},$$

and covariance

$$\Sigma_{12} = \text{Cov}(y_t, y_{t-1}) = \frac{\pi_0 \pi_1 (\mu_0 - \mu_1)^2}{1 - \phi_1^2} \left[ \phi_1 (1 + \phi_1^2) - 2\phi_1 (p_{0|0} + p_{1|1} - 1) \right] + \frac{\phi_1 \sigma_e^2}{1 - \phi_1^2}.$$
Bibliography


Bibliography


Bibliography


Gillemot, L., Farmer, J. and Lillo, F.: 2006, There’s more to volatility than volume, Quantitative Finance, 25. doi:10.1080/14697680600835688. [38]


Bibliography


Bibliography


Neyman, J. and Scott, E. L.: 1948, Consistent Estimates Based on Partially Consistent Observations, Econometrica 16(1), 1–32. [56, 57]


