

APPLICATIONS OF STOCHASTIC CONTROL AND STATISTICAL INFERENCE IN MACROECONOMICS AND HIGH-DIMENSIONAL DATA

A Thesis
Presented to
The Academic Faculty

by

Zhi Han

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2015

Copyright © 2015 by Zhi Han

APPLICATIONS OF STOCHASTIC CONTROL AND STATISTICAL INFERENCE IN MACROECONOMICS AND HIGH-DIMENSIONAL DATA

Approved by:

Professor Shijie Deng, Advisor
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Professor Xiaoming Huo, Advisor
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Professor Ton Dieker
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Professor Enlu Zhou
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Professor Liang Peng
Robinson College of Business
Georgia State University

Date Approved: 07 Aug 2015

To the people I love.

ACKNOWLEDGEMENTS

I want to thank my advisors, Dr. Shijie Deng and Dr. Xiaoming Huo for guiding and supporting me over the past four years. Without their help, I would not have made so much progress and completed my thesis.

I would like to thank my thesis committee members, Dr. Ton Dieker, Dr. Enlu Zhou, and Dr. Peng Liang, for their guidance and valuable comments throughout this whole process.

I would like to thank my friends, classmates, and colleagues for their support during my PhD study. We really had a great time in Atlanta over the past five years.

Finally, I would like to thank my parents and my wife for their unconditional love. Without their support, I would not have had the chance to pursue and obtain my doctoral degree.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
SUMMARY	x
I OVERVIEW	1
1.1 Management of foreign exchange reserves	1
1.1.1 Background	1
1.1.2 Contributions of the work	2
1.2 Fast computing algorithm of partial distance covariance	3
1.2.1 Background	3
1.2.2 Contributions of the work	4
1.3 High dimension feature screening using partial distance covariance	5
1.3.1 background	5
1.3.2 Contributions of the work	6
II A STOCHASTIC CONTROL MODEL FOR FOREIGN EX- CHANGE RESERVE MANAGEMENT	8
2.1 Introduction	8
2.2 Empirical Evidence	12
2.3 Brownian drift control model	15
2.4 Verification theorem	18
2.5 Explicit solution in the two drift case	21
2.5.1 Control band policies	21
2.6 Optimal policy and optimal parameters	28
2.7 Numerical Results	32
2.7.1 Optimal Cost Comparison	32

2.7.2	Sensitivity Analysis	35
2.8	Explicit solution in the three drifts case	37
2.8.1	Control band policies	38
2.8.2	Optimal policy and optimal parameters	43
2.8.3	Numerical Results	47
2.9	Algorithm for an arbitrary number of drifts	50
2.9.1	Control band policies	52
2.9.2	Optimal policy and optimal parameters	55
III PARTIAL DISTANCE COVARIANCE: A FAST COMPUTING ALGORITHM		63
3.1	Introduction	63
3.2	Partial distance covariance	67
3.2.1	Sample pdCov and pdCor	68
3.3	An $O(n \log n)$ algorithm	70
3.4	Numerical Results	72
3.5	Algorithms	73
IV APPLICATION OF PARTIAL DISTANCE COVARIANCE IN HIGH DIMENSIONAL FEATURE SCREENING		80
4.1	Introduction	80
4.2	Some preliminaries	82
4.2.1	Partial distance covariance	82
4.3	An Iterative Screening Procedure	84
4.4	Simulations Results	86
4.5	Real application on the Leukemia data	91
4.5.1	Data description	91
4.5.2	Partial distance covariance gene selection results based on all the observations	92
4.5.3	Fivefold cross validation	94
APPENDIX A — A STOCHASTIC CONTROL MODEL FOR FOREIGN EXCHANGE RESERVE MANAGEMENT		97

APPENDIX B — PARTIAL DISTANCE COVARIANCE: A FAST COMPUTING ALGORITHM	104
REFERENCES	108

LIST OF TABLES

1	Dependency on h	37
2	Dependency on U	37
3	Dependency on $K(\mu_1, \mu_2)$	37
4	Dependency on $K(\mu_2, \mu_1)$	38
5	Dependency on h	50
6	Dependency on U	50
7	Dependency on $K(\mu_1, \mu_2)$	51
8	Dependency on $K(\mu_2, \mu_3)$	51
9	Running times (in seconds) for the direct and fast methods for computing the partial distance covariances. The values in the parentheses are sample standard errors. At each sample size, 1,000 repetitions were run.	73
10	Running times (in seconds) for the fast methods for computing the partial distance covariances. The values in the parentheses are sample standard errors. At each sample size, 1,00 repetitions were run. . . .	74
11	Results of simulated example I: accuracy of SIS, ISIS, and PDCOV in including the true model $\{X_1, X_2, X_3, X_4\}$	88
12	Results of simulated example I: average number of iterations in ISIS and PDCOV are taken to include the true model $\{X_1, X_2, X_3, X_4\}$. .	88
13	The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum number of iterations \mathcal{L} out of 500 replications in Example 2	90
14	Partial distance covariance gene selection results based on all the observations. The genes in bold type are also among the 50 selected genes in [12].	93
15	The optimal tuning parameters with corresponding prediction errors .	95

LIST OF FIGURES

1	Reserves in Individual Countries from 1985 to 2001. Note: The vertical axis measures reserves in billions of dollars; the horizon axis is the year. Data source www.theglobaleconomy.com	13
2	The continuation and intervention regions in the two drift case	23
3	The difference between $c^*(\beta)$ and $V_1(a_m)$ for different parameters $h, U, K(\mu_1, \mu_2)$, and $K(\mu_2, \mu_1)$ in two drift case	33
4	The continuation and intervention regions in the three drift case . . .	39
5	The difference between $M^*(\beta)$ and $V_1(x_{21}^m)$ for different parameters $h, U, K(\mu_1, \mu_2)$, and $K(\mu_2, \mu_3)$ in three drift case	49
6	A comparison of running time between the direct and fast methods for the computation of the partial distance covariances	74
7	An illustration of running times of the fast method for the computation of the distance correlations. The red solid line corresponds to an $O(n \log n)$ algorithm.	75
8	Dynamic of the reserve accumulation in three drift threshold stopping rule	98

SUMMARY

This thesis is focused on the optimality of stochastic control in macroeconomics and the fast algorithm of statistical inference. The first topic involves the modeling of foreign exchange reserve management as an optimal drift control problem. The second topic involves the fast computing algorithm of partial distance covariance statistics with its application in feature screening in high dimensional data.

In the first part of the dissertation, we study the problem of optimally controlling the level of foreign exchange reserves held by a country. When a reserve authority accumulates foreign exchange reserves to meet changing economic conditions, it faces the challenge of finding the right balance between the holding costs and the operational costs involved in adjusting the reserve size. We consider a foreign exchange reserve whose inventory fluctuation is modeled by a Brownian motion with drift, and at any moment the reserve manager can adjust the inventory level by varying the drift at which the reserve accumulates or depletes, but incurs a cost which satisfies triangle inequality. When the reserve is accumulating or depleting, it also incurs a maintaining cost related to the current drift. The inventory level must be nonnegative at all times and continuously incur a linear holding cost. The reserve manager's problem is to decide when and how to change the drift so that the long run expected discounted cost of maintaining the foreign exchange reserve is minimized. We show that, under certain conditions, the control band policies are optimal for the discounted cost drift control problem and explicitly calculate the parameters of the optimal control band policy. In the two drift case, this form of policy is described by two parameters $\{L, U\}$, $0 < L < U$. When the inventory falls to L (rises to U), the controller switches the drift rate to depletion (accumulation). We also extend the result to the multiple

drift case and develop an algorithm to calculate the optimal thresholds of the optimal control band policy.

In the second part of the dissertation we study the problem of fast computing algorithm of partial distance covariance. If the computation of partial distance covariance is implemented directly accordingly to its definition then its computational complexity is $O(n^2)$ which may hinder the application of an algorithm. To illustrate it, if n is equal to 10^6 , an $O(n^2)$ algorithm would need 10^{12} numerical operations, which is impossible even for modern computers. In comparison, an $O(n \log n)$ algorithm would only require around 10^6 numerical operations, which is doable. In this part of the thesis, we show that an $O(n \log n)$ algorithm for a version of the partial distance covariance exists. The derivation of the fast algorithm involves significant reformulation of the original version of partial distance covariance. We also demonstrate its application in feature screening in high dimensional data in the following part of the thesis.

In the final part of the thesis we further study the feature screening problem in high dimensional data. We propose an iterative feature screening procedure based on the partial distance covariance. This procedure can simultaneously address the following two issues when using sure independence screening (SIS) procedure. First, an important predictor that is marginally uncorrelated but jointly correlated with the response cannot be picked by SIS and thus not entering the estimation model. Second, SIS works only for linear models, and performance is very unstable in other nonlinear models. To the best of our knowledge, this is the first time that a “new metric” – partial distance covariance – is used for feature screening in high dimensional data, and the idea of conditional screening is formally developed.

CHAPTER I

OVERVIEW

1.1 Management of foreign exchange reserves

1.1.1 Background

Foreign exchange reserves are assets held by central banks and monetary authorities, usually in different reserve currencies, mostly the United States dollar, and used to back its liabilities. These reserve assets allow a central bank to purchase domestic currency, which is considered a liability for the central bank. Thus, the quantity of foreign exchange reserves can change as a central bank implements monetary policy.

The cost in holding large foreign exchange reserves consists of two parts: The first part is the maintaining cost, as fluctuations in exchange markets result in gains and losses in the purchasing power of reserves. In addition to fluctuations in exchange rates, the purchasing power of fiat money decreases constantly due to devaluation through inflation. Furthermore, large currency reserves could have been invested in higher yielding assets, which is the opportunity cost. The second part is the adjustment cost, because any reserve accumulation change as a consequence of monetary policy change can make an impact on the macro economic.

The massive accumulation of foreign reserves by many countries has challenged the authorities regarding prudent reserve management. One strand of research adopted by researchers is to apply impulse control. It describes the situation in which an injection of reserves, perhaps from some institutions like the World Bank or the International Monetary Fund (IMF), immediately restores the reserves to their target level when reserves hit a lower bound. While this model is straightforward, it still has some shortcomings. First, the pattern implied by the impulse control is not supported

by the real data. Actually, the reserve patterns of many countries give evidence of gradual decline and increase that are bounded from below and above, rather than the shark-tooth pattern implied by the impulse control model. Second, in real life, a more common method for reserve managers is to change monetary policy. This change alters the financial flow which gradually increases or decreases the reserve level.

In the first part of the dissertation, we apply a drift control model to foreign exchange reserves management. The key feature of the drift control model is that the reserve managers control the upward and downward drift rates of the reserve. The objective is to minimize the cost subject to the constraint that reserve managers can choose when and how to change the drift of the reserves. We model the reserve process in the absence of any control, as a Brownian Motion with drift μ and variance σ^2 . The reserve continuously incurs linear holding costs and maintenance costs associated with the reserve drift, and it must remain non-negative at all times. The reserve managers may, at any time, adjust the drift by, for example, executing different monetary policies or intervening in the foreign exchange market, but incur a cost for changing the drift that satisfies the triangle inequality.

1.1.2 Contributions of the work

In Chapter 2, we show that, under certain conditions, a simple form of policy, called *Control Band Policy*, is optimal for the drift control problem in the two drifts case. This form of policy is described by two parameters $\{L, U\}$, $0 < L < U$. When the reserve level rises to U (falls to L), the controllers switch the drift to depletion (accumulation). Control band policy can be interpreted as maintaining the same reserve drift each time and adjusting this quantity only when the reserve level is either too high or too low. We also extend the result to the multiple drifts case and develop an algorithm to calculate the optimal thresholds of the optimal control band

policy.

1.2 Fast computing algorithm of partial distance covariance

1.2.1 Background

Distance covariance and distance correlation are scalar coefficients that characterize independence of random vectors in arbitrary dimension. They were introduced in 2005 by G. J. Székely, in several lectures to address the deficiency of Pearson's correlation, namely that the Pearson's correlation can be zero for dependent variables.

The distance covariance, denoted $\mathcal{V}(X, Y)$, of two random vectors X and Y characterizes independence; that is

$$\mathcal{V}(X, Y) \geq 0$$

with equality to zero if and only if X and Y are independent. This coefficient is defined by L_2 norm measuring the distance between the joint characteristic function (c.f.) $\phi_{X,Y}$ of X and Y , and the product $\phi_X\phi_Y$ of the marginal c.f.'s of X and Y .

Properties, extensions and applications of distance covariance and distance correlation have been discussed in the recent literature; see, for example, [16] and [18]. A natural question then arises, “How do we define partial distance covariance (correlation) which extends distance covariance (correlation) in a similar sense that partial correlation extends correlation?” One could try to follow the definitions of the classical partial covariance and partial correlation that are based on orthogonal projections in a Euclidean space, but there is a serious difficulty. Orthogonality in the case of partial distance covariance and partial distance correlation means independence, but when we compute the orthogonal projection of a random variable onto the condition variable, the “remainder” in the difference is typically not independent of the condition.

Alternately, the form of sample distance covariance (Definition 3.1.2) may suggest

an inner product, so one might think of working in the Hilbert space of double centered distance matrices defined as equation (116) in section 3.1, where the inner product is the squared distance covariance statistic defined as equation (118) in the same section. Here, we are facing another problem: what would the projections represent? The difference D of double centered distance matrices is typically not a double centered distance matrix of any sample. This does not affect formal computations, but if we cannot interpret our formulas in terms of samples then inference becomes impossible.

To overcome these difficulties while preserving the essential properties of distance covariance, [27] finally arrived at an elegant solution which starts with defining an alternate type of double centering called “ \mathcal{U} -centering” (see Definition 3.2.1 and Proposition 3.2.2). The corresponding inner product is an unbiased estimator of squared population distance covariance. In the Hilbert space of \mathcal{U} -centered matrices, all linear combinations, and in particular projections, are zero diagonal \mathcal{U} -centered matrices.

As a newly developed concept, partial distance covariance has the advantage that it can capture nonlinear dependence [27]. It is expected that partial distance covariance has application in the broad field of life science, engineering, and finance. Particularly as an example, we have successfully applied partial distance covariance to feature screening in high dimensional data, see Chapter 4. If partial distance covariance was implemented directly as it is defined, its computational complexity can be as high as a constant times n^2 for a sample size n . This fact has been regarded as a disadvantage of adopting partial distance covariance.

1.2.2 Contributions of the work

In Chapter 3, we demonstrate that an $O(n \log n)$ algorithm for a version of the partial distance covariance exists. To illustrate how an $O(n^2)$ order of complexity may hinder the application of an algorithm, assume that n is equal to 10^6 . An $O(n^2)$ algorithm would need 10^{12} numerical operations, which is impossible even for modern computers.

In comparison, an $O(n \log n)$ algorithm would only require around 10^6 numerical operations, which is doable. The main idea behind the proposed new algorithm is to use a technique rooted in the the AVL tree structure [1]. The same idea has been utilized to develop a fast algorithm for computing the Kendalls τ rank correlation coefficient ([15]; [6]). We extend it to make it suitable for our purpose. The derivation of the fast algorithm also involves significant reformulation of the original version of partial distance covariance.

1.3 High dimension feature screening using partial distance covariance

1.3.1 background

The idea of feature screening came along as high-dimensional data were collected in modern technology. It was aimed at dealing with the challenges of computational expediency, statistical accuracy, and algorithmic stability because of high dimensionality. [10] proposed the sure independence screening (SIS) and showed that the Pearson correlation ranking procedure possessed a sure screening property for linear regression with Gaussian predictors and responses. However, two potential issues might arise with the screening procedures. First, an important predictor that is marginally uncorrelated but jointly correlated with the response cannot be picked by SIS and thus will not enter the estimation model. Second, this procedure works only for linear models, and performance is very unstable in other nonlinear models.

To address the first issue, an iterative SIS (ISIS) was proposed in [10] as an extension of SIS. The ISIS works as follows: In the first step, we select a subset of variables using an SIS-based model selection model. Then we have a vector of residuals from regressing the response over the variables selected in the first step. In the next step, we treat those residuals as the new response and utilize the same method as in the previous step to the remaining variables. From the discussion above, the ISIS uses a

residual-based approach to circumvent the problem but the idea of conditional screening has never been formally developed. However, the performance of ISIS in nonlinear models still remains unknown. To address the second issue, a new feature screening procedure for high-dimensional data based on distance correlation, named DC-SIS, was presented in [18]. DC-SIS retained the sure screening property of the SIS, and because distance correlation was applicable to arbitrary distributions, DC-SIS could also be used for screening features without specifying a regression model between the response and the predictors and thus was robust to model mis-specification. However, similar to the SIS, the DC-SIS may fail to identify some important predictors that are marginally independent of the response.

In the third part of the dissertation, we propose an iterative feature screening procedure based on partial distance covariance. In the initial step of variable selection, the first variable to enter the model is the variable x_j for which distance covariance $C_{x_j, y}$ with response y is largest. After the initial step, we have a model with one predictor x_j , and we compute partial distance covariance of (y, x_k) conditional on x_j , for the variables $x_k \neq x_j$ not in the model, then select the variable x_k for which partial distance covariance is largest. Then continue, at each step computing partial distance covariance of (y, x_j) conditional on w for every variable x_j not yet in the model, where w is the vector of predictors currently in the model. The variable to enter next is the one that maximizes partial distance covariance. The stopping rule is set at 5% significant level for zero partial distance covariance coefficient test.

1.3.2 Contributions of the work

In Chapter 4, we propose an iterative feature screening procedure based on partial distance covariance. We show that $pdcov(x, y; z)$ equals to the distance covariance of U and y , where U is a random vector such that the \mathcal{U} -centered distance matrix of its sample is exactly equal to $P_{z^\perp}(x)$ (124), and U is independent of z . Therefore,

after selected variables enter the model, those that are marginally weakly correlated with response purely due to the presence of variables in conditional set should now be correlated with the response. This addresses the first issue that SIS may miss an important predictor that is marginally uncorrelated but jointly correlated with the response. Partial distance covariance also has the advantage of capturing nonlinear dependence. It can be used for selecting variables without model specification. This helps to solve the second issue that SIS may fail in nonlinear models. To the best of our knowledge, this is the first time that a “new metric” – partial distance covariance – is used for feature screening, and the idea of conditional screening is formally developed. At the end, we demonstrate the performance of our procedure through simulations and a real example.

CHAPTER II

A STOCHASTIC CONTROL MODEL FOR FOREIGN EXCHANGE RESERVE MANAGEMENT

2.1 Introduction

Foreign exchange reserves are assets held by central banks and monetary authorities, usually in different reserve currencies, mostly the United States dollar, that are used to back its liabilities. These reserve assets allow a central bank to purchase the domestic currency, which is considered a liability for the central bank. Thus, the quantity of foreign exchange reserves can change as a central bank implements monetary policy.

The costs in holding large foreign exchange reserves consists of two parts: The first part is the maintaining cost, as fluctuations in exchange markets result in gains and losses in the purchasing power of reserves. In addition to fluctuations in exchange rates, the purchasing power of fiat money decreases constantly due to devaluation through inflation. Furthermore, large currency reserves could have been invested in higher yielding assets, which is the opportunity cost. The second part is the intervention cost. Central banks or monetary authorities may approach IMF for bail-out when countries are in deep financial crises or their reserve levels fall below precautionary level. Furthermore, any reserves accumulation change as a consequence of monetary policy adjustment will influence the cost of credit and debt.

The massive accumulation of foreign exchange reserves by many countries has challenged the authorities about prudent reserve management due to the significant role of foreign exchange reserves in macro economics. One strand of research adopted by researchers is to apply impulse control. It describes the situation in which an injection of reserves, perhaps from some institutions like the World Bank or IMF,

immediately restores the reserves to their target level when reserves hit a lower bound. While this model is straightforward, it still has some shortcomings. First, the pattern implied by the impulse control model is not supported by the real data. Figure 1 illustrates the reserve patterns for four countries over the 1985-2001 period. Observe that reserves are not restocked immediately after hitting a lower threshold. In fact, the reserve patterns show that reserve levels gradually increase and decrease.

Among all the popular stochastic control models, drift control is a widely adopted one, in which the controller may change the drift when the diffusion process reaches a threshold. The drift control model has been widely applied to problems of inventories, queuing systems, and economics. [5] studied the problem of minimizing the infinite horizon expected average cost of a one side reflected diffusion process in which the controller can switch between two sets of drifts and volatility parameters. The problem involves operating, switching, and holding cost. The author determined the optimal policy via the dynamic programming. [19] extends this problem to a more general setting in which we can select from multiple drifts, and they developed a novel solution approach based on linear programming. [2] applied the drift control model to minimize the infinite horizon expected discounted cost of managing international reserves. They assume that the reserve authority chooses between two drifts: the upward and downward drifts of the reserve. However, they take the form of the policy as given and calculate the cost based on martingale stopping theory, then they derive the optimal policy numerically.

Second, the impulse control is not the normal way that reserve authorities control the reserve level. According to [23], the main drives for reserve accumulation of the emerging market economies are self-insurance against financial crises and the pursuit of export-led growth supported by exchange rates anchored de jure or de facto to the US dollar. So the reserve authorities change monetary or exchange-rate policy on their own initiative to alter financial account flows that gradually increase or decrease

the level of reserves. From the modeling point of view, the policy change alters the drift of reserves accumulation, rather than restock reserves to their target level in the abrupt manner characterized by the impulse control model.

We model the problem of managing demand for foreign exchange reserves as a Brownian drift control problem and seek a policy that minimizes the long-run discounted cost. The key feature of the drift control model is that the reserve managers control the drift of the reserve. We model the reserve process in the absence of any control, as a Brownian Motion with drift μ and variance σ^2 . Inventory continuously incurs linear holding cost and maintaining cost associated with the reserve drift, and it must remain non-negative at all times. According to [24], the holding cost of reserves equals to the spread between the private sector's cost of short-term borrowing abroad and the yield that the Central Bank earns on its liquid foreign assets. This validates the assumption of linear holding cost. We also assume the reserve authority can, at some cost, shift the rates of reserve accumulation and depletion among a finite set of alternatives, for example by choosing different exchange-rate or monetary/fiscal policies. We model the cost of changing the accumulation rate from u to v as a fixed cost $K(u, v) > 0$ and assume the cost function $K(\cdot, \cdot)$ satisfies the usual triangle inequality so that changing the processing rate from u to v in a single step is no more expensive than accomplishing the same change via a series of intermediate steps. Even with higher accumulation rate, a series of bad financial shocks can push the reserve level below zero. To ensure that the reserves remains nonnegative, we impose instantaneous controls at the lower boundary, corresponding roughly to obtaining additional reserves from the IMF or other countries, for instance, and this creates the intervention cost.

Our model differs from the ones in the above cited works in many ways. While [5] and [19] address the long-term average cost problem, we study the drift control problem in the discounted cost case, as a discount factor is natural and intuitive in

financial applications. While [2] studied the drift control problem in the discounted cost case, they assumed that the reserve authority chooses between two drifts: the upward and downward drifts of the reserves; however, they took the form of the control band policy as given and calculated the cost based on martingale stopping theory, then they derive the optimal policy numerically. We, on the other hand, extend the control policy space to all the non-anticipated policy, and prove that under certain conditions, the control band policy is optimal among all non-anticipated policy. Our empirical results show that the value function derived from our optimal stopping theory has superior performance to the value function derived from the martingale approach in [2]. Furthermore, we extend the drift control problem with only two drifts to the case in which we can switch from arbitrarily finite drifts, and develop an algorithm to calculate the optimal thresholds of the optimal control band policy. The drift control problem with only two drifts, the upward and downward drifts of the reserve, is too simple to characterize the real dynamics of the reserves accumulation. Normally, the central bank can adopt at least three main types of fiscal policy to change the reserve accumulation: neutral fiscal policy, expansionary fiscal policy, and contractionary fiscal policy. Therefore, our work is more general and more valuable in decision making for reserve authorities.

The rest of the chapter is organized as follows. In Section 2.2, we present empirical evidence on country reserve holdings to motivate the consideration of a drift control model. In Section 2.3, we describe the discounted cost Brownian drift control problem and its policy space. Section 2.4 sets up the preliminaries for the solution approach to the problem. In Section 2.4 we establish a lower bound for the optimal expected total discounted cost and prove the “Verification Theorem”. In Section 2.5 we define a value function for control band policies with discounted cost criteria in the two-drift case, and show that the expected total discounted cost can be calculated through this function. In Section 2.6, we first find a control band policy which is optimal among

all control band policies. We then prove that, under certain conditions, the control band policy is optimal for the discounted cost Brownian drift control problem among all non-anticipated policies by applying the verification theorem, and derive explicit equations for calculating the optimal control policy parameters. In Section 2.7, we provide numerical examples illustrating the optimality of the drift control policy. We extend the result to the drift control problem with three drifts case in Section 2.8 and develop an algorithm for finding the optimal control band policy for the case with an arbitrary number of drifts in Section 2.9.

2.2 Empirical Evidence

To motivate the modeling of the foreign exchange reserve management as a drift control problem, we present empirical evidence which illustrates the association between changes in the direction of reserve drift and explicit policy-change decisions made by the reserve authority to switch the drift. To document the dynamics of international reserves, we examine the empirical properties of monthly international reserves for 90 countries over the sixteen-year period 1985 - 2001. Figure 1 illustrates the reserve dynamics for a representative set of these countries.

To document evidence of policy changes and turning points in international reserve holdings, we use a case study approach. We find the policy changes are motivated primarily or in part by concerns about reserve levels, and the switch from different drifts in reserves may also be triggered by an external shock rather than an explicit policy change. In the aftermath of recent financial crises, some countries appear to have extended the reserve holding period with upward drift in their desire to accumulate reserves. To appreciate the nature of this evidence, we present some case studies drawn from countries highlighted in Figure 1.

South Africa: South Africa began to use forward market to encourage and facilitate the use of foreign trade credits by domestic corporates as a mechanism for

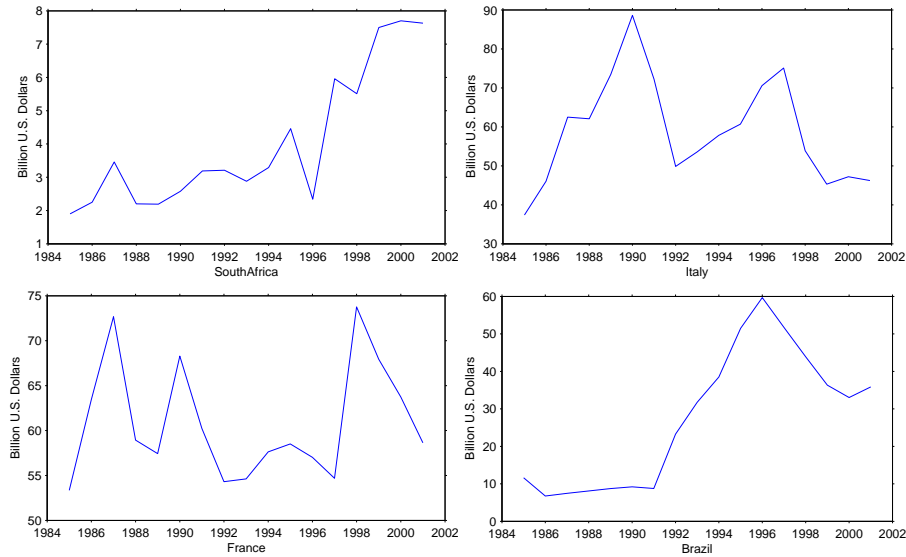


Figure 1: Reserves in Individual Countries from 1985 to 2001. Note: The vertical axis measures reserves in billions of dollars; the horizon axis is the year. Data source www.theglobaleconomy.com.

its Central Bank to accumulate foreign exchange reserves in September 1985. Reserves continued to drift upwards slowly until reaching a local peak in March 1995, interrupted by several sharp drops in reserves associated with contagion from the emerging markets crisis. After the crisis, however, capital flight from South Africa continued and reserves fell further. Worried about its reserve position, South Africa started accumulating reserves by buying foreign exchange reserves on a spot basis after reserves reached a local trough in May 1998.

Italy: A few months before the first reserve peak in August 1990, the Italian government cited Italy's strong balance of payments position and sizable reserve holdings to justify its decision to adjust the lira's value and adopt the narrower 2.25% band of fluctuation used by fellow members in the European Exchange-Rate Mechanism (ERM). Just after the reserve trough in August 1992, Italy expressed concern about declining reserves in the face of strong speculation against the lira. It dropped out of the ERM and let the lira float. In the months prior to the December 1997 reserve peak, the Italian authorities noted that a stronger lira and stronger reserve position

had loosened the constraints on monetary and exchange-rate policies. Italy rejoined the ERM and started a period of interest-rate reductions, including one in the month when reserves reached their local peak. Reserves then started to fall, reaching a new low at the start of 1999 when Italy joined the European Monetary Union (EMU).

France: In the years 1985–2001, France’s reserve holdings reached local peaks in 1987 and 1998, and local troughs in 1989, 1993, and 1997. In the years 1983-85 just before the first peak, the French authorities decided to move towards a strategy geared to lowering inflation and to increasing the competitive position of the economy. In the 1993 reserve trout, French authorities confirmed the commitment to a stable exchange rate on August 2nd, with the decision to maintain unchanged the central rate of the franc within a fluctuation band enlarged to 15%,after strong tensions which affected the exchange markets notably in 1992-1993. Reserves then started to fall on December 3, 1998, after French authorities decided to bring down the main rate to 3% within a coordinated movement of lowering interest rates.

Brazil: After a period of low and stable reserves, the reserve holdings of Brazil began to rise dramatically as a result of fixed exchange rate policy starting from the fall of 1991. Reserves continued to drift upwards until the end of 1997, interrupted by three sharp drops in reserves associated with contagion from external financial crises—the Mexican crisis (December 1994), the Asian crisis (July 1997), and the Russian crisis (August 1998). After all three financial crises, Brazil reacted by increasing interest rates sharply to reduce capital flight and improve its reserve position. However, capital flight from Brazil continued and reserves fell further, causing the authorities to seek a large support package from the IMF in the fall of 1998, and to increase the exchange-rate band on January 13, 1999. Since 2001, Brazilian reserves have trended upward, aided by periodic increases in domestic interest rates and external financing extended by the IMF to protect Brazil from adverse spillovers from Argentina’s crisis

To summarize, the exchange reserve patterns of several countries have strongly

supported a drift control model that associates changes in the direction of reserve drift with explicit decisions made by the reserve authority to switch the drift by changing monetary policy. A model that allows for finite periods of reserve accumulation and depletion and finite drift rates in both directions also finds support in the data.

2.3 *Brownian drift control model*

Let $X = \{X(t), t \geq 0\}$ be a Brownian motion with drift $\mu(t)$ in some fixed finite set Λ for each $t \geq 0$, variance $\sigma^2(t)$, starting from x . Then X has the following representation

$$X(t) = X(0) + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dB(s), \quad t \geq 0, \quad (1)$$

where $B = \{B(t), t \geq 0\}$ is a standard Brownian motion that has drift 0, variance 1, starting from 0. We assume B is defined on some filtered probability space $(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, \mathbb{P})$ and B is an $\{\mathcal{F}_t\}$ -martingale. Thus, B is also known as an $\{\mathcal{F}_t\}$ -adapted standard Brownian motion. To simplify notation, we consider the special case in which $\sigma(t) = \sigma$. The process $X(t)$ describes the difference between cumulative input flow and cumulative output flow of the foreign exchange reserves by time t , i.e., the “netput” process. The drift $\{\mu(t), t \geq 0\}$, which is adapted to the Brownian motion $\{B(t), t \geq 0\}$, is the average accumulation rate of the reserves. We assume that the monetary authorities can, at some cost, shift the accumulation rate among a finite set of alternatives by changing the monetary policy.

Furthermore, let $A(t)$ denote the minimal amount of regulation (reserves injection from the world bank or IMF) necessary to keep the reserve level from falling below the boundary 0 up to time t , where $A(t)$ is defined as

$$A(t) = -\min\{0, \min_{s \leq t} X(s)\}.$$

The regulation $A(t)$ satisfies

- (1) $A(t)$ are adapted to the filtration $\{\mathcal{F}_t\}$.

(2) $A(t)$ are right continuous, increasing, and non-negative.

The controlled process is

$$R(t) = R(0) + \int_0^t \mu(s) ds + \sigma B(t) + A(t), \quad t \geq 0 \quad (2)$$

where $R(0) = X(0)$. We associated with $A(t)$ the controlled process $R(t) = X(t) + A(t)$, and $A(t)$ is said to be *feasible* if

$$\mathbb{P}_x(R(t) \geq 0 \text{ for all } t \geq 0) = 1, \quad \forall x \geq 0. \quad (3)$$

$$\mathbb{E}_x \left[\int_0^\infty e^{-\beta t} dA(t) \right] < \infty, \quad \forall x \geq 0. \quad (4)$$

where β denotes the discounted rate, and $\mathbb{P}_x(*) = \mathbb{P}(*/R(0) = x)$, $\mathbb{E}_x[*] = \mathbb{E}[*/R(0) = x]$.

The controlled process $R(t)$ lives in the positive region $[0, \infty)$, and the controller may adjust the drift by choosing them from the values in the finite set Λ . Let $\mathcal{D} = \{(a, \lambda) : a \in [0, \infty), \lambda \in \Lambda\}$ denote the domain of the controlled process.

We restrict our attention to the space \mathcal{P} of all non-anticipated policy

$$\Phi = \{(\tau_n, u_n) : n \geq 0\} \quad (5)$$

where

- (i) $0 = \tau_0 < \tau_1 < \dots$ is a sequence of stopping times,
- (ii) each u_n is a random variable adapted to \mathcal{F}_{τ_n} with value in Λ indicating the rate to which we change the drift at time τ_n .

Under the policy $\Phi = \{(\tau_n, u_n) : n \geq 0\}$, the drift has the value $\mu(t) = u_n$ for $\tau_n \leq t < \tau_{n+1}$.

To change the drift from μ_i to μ_j , the controller must pay a fixed cost $K(\mu_i, \mu_j) > 0$, for $\mu_i \neq \mu_j$, which satisfies a triangle inequality: $K(\mu_i, \mu_j) + K(\mu_j, \mu_k) \geq K(\mu_i, \mu_k)$ for all drifts μ_i, μ_j , and μ_k .

There is a cost $c_i = c(\mu_i)$ per unit time for maintaining the drift μ_i . Let h be the holding cost of \$1 per unit time and we assume there is cost U per \$1 of regulation at the boundary 0. According to [24], the holding cost of reserves is equal to the spread between the private sector's cost of short-term borrowing abroad and the yield that the Central Bank earns on its liquid foreign assets. This validates the assumption of linear holding cost

We consider the following *discounted cost Brownian control problem*, which is to obtain a non-anticipated policy that minimizes the expected discounted cost in infinite horizon :

$$\begin{aligned} V_i(x) &= \min_{\Phi \in \mathcal{P}} DC(x, \mu_i, \Phi) \\ &= \min_{\Phi \in \mathcal{P}} \mathbb{E}_{x,i} \left[\int_0^\infty e^{-\beta t} (c(\mu(t)) + hR(t)) dt + \int_0^\infty e^{-\beta t} U dA(t) + \sum_{n=1}^\infty e^{-\beta \tau_n} K(u_{n-1}, u_n) \right]. \end{aligned} \quad (6)$$

where $(R(0), u_0) = (x, \mu_i)$.

Note that when considering a function f defined on a subset of $\mathbb{R} \times \Lambda$, we treat f as a family of functions $\{f(\cdot, \mu_i) : \mu_i \in \Lambda, i = 1, 2, \dots, k\}$, each defined on the corresponding subset of \mathbb{R} ; so for example, use f' and f'' to represent derivatives with respect to the first argument. For a function f we define the following quasi-variational inequalities.

Definition 2.3.1. *We say that functions $f(\cdot, \mu_i) : [0, \infty) \rightarrow [0, \infty), i = 1, 2, \dots, k$ satisfy the quasi-variational inequalities (QVI) for problem in equation (6) if for each $i = 1, 2, \dots, k$*

$$\inf_{x \in [0, \infty)} \{ \mathcal{L}_i f(x, \mu_i) + c(\mu_i) + h \cdot x, \quad \min_{j \neq i} \{ f(x, \mu_j) + K(\mu_i, \mu_j) \} - f(x, \mu_i) \} = 0, \quad (7)$$

where

$$\mathcal{L}_i w(x) = \frac{1}{2} \sigma^2 w''(x) + \mu_i w'(x) - \beta w(x)$$

We observe that a solution $f(\cdot, \mu_i)$ of the QVI separates the interval $[0, \infty)$ into two disjoint regions: a continuation region

$$\mathcal{C}_i := \{x \in [0, \infty) : \min_{j \neq i} \{f(x, \mu_j) + K(\mu_i, \mu_j)\} > f(x, \mu_i), \mathcal{L}f(x, \mu_i) + c(\mu_i) + h \cdot x = 0\}$$

and an intervention region

$$\mathcal{S}_i := \{x \in [0, \infty) : \min_{j \neq i} \{f(x, \mu_j) + K(\mu_i, \mu_j)\} = f(x, \mu_i)\}$$

Definition 2.3.2. *The following non-anticipated policy is called the QVI-control associated with $\{f(\cdot, \mu_i), i = 1, \dots, k\}$:*

$$\tau_1 = \inf\{t > \tau_0 : f(R(t), u_0) = \min_{\mu_j \neq u_0} \{f(R(t), \mu_j) + K(u_0, \mu_j)\}\}, \quad (8)$$

$$u_1 = \arg \min_{\mu_j \neq u_0} \{f(R(t), \mu_j) + K(u_0, \mu_j)\}, \quad (9)$$

and, for $n \geq 2$:

$$\tau_n = \inf\{t > \tau_{n-1} : f(R(t), u_{n-1}) = \min_{\mu_j \neq u_{n-1}} \{f(R(t), \mu_j) + K(u_{n-1}, \mu_j)\}\}, \quad (10)$$

$$u_n = \arg \min_{\mu_j \neq u_{n-1}} \{f(R(t), \mu_j) + K(u_{n-1}, \mu_j)\}. \quad (11)$$

2.4 Verification theorem

In this section, we state and prove a theorem that establishes a lower bound for the optimal expected total discounted cost. This theorem is also called the “Verification Theorem” in the literature.

Theorem 2.4.1. *Suppose that for each $\mu_i \in \Lambda$, $f(\cdot, \mu_i) \in C^1([0, \infty); [0, \infty)) \cap C^2([0, \infty) \setminus \mathcal{N}; [0, \infty))$ where \mathcal{N} is a finite subset of $[0, \infty)$, is a solution of the QVI, and there exists a constant $M > 0$ such that $|f'(x, \mu_i)| < M$ for all $x \in [0, \infty)$. Assume further that*

$$f'(0, \mu_i) \geq -U, \quad i = 1, 2, \dots, k. \quad (12)$$

Then, for every $\Phi \in \mathcal{P}$ with initial state $(R(0), u_0) = (x, \mu_i) \in \mathcal{D}$, we have

$$f(x, \mu_i) \leq DC(x, \mu_i, \Phi). \quad (13)$$

Furthermore, if the QVI-control Φ^* associated with $\{f(\cdot, \mu_i), i = 1, \dots, k\}$ is admissible and

$$f'(0, \mu_i) = -U, \quad \text{if } 0 \in \mathcal{C}_i \quad i = 1, 2, \dots, k,$$

then Φ^* is the optimal drift control policy, and for every initial state $(R(0), u_0) = (x, \mu_i) \in \mathcal{D}$, we have

$$f(x, \mu_i) = V_i(x) = DC(x, \mu_i, \Phi^*). \quad (14)$$

Proof. For every $T > 0$ and every policy Φ with initial state $(R(0), u_0) = (x, \mu_i)$, we can write

$$\begin{aligned} & e^{-\beta T} f(R(T), \mu(T)) - f(x, \mu_i) \\ &= \sum_{j=1}^{N(T)} \{e^{-\beta \tau_j} f(R(\tau_j-), u_{j-1}) - e^{-\beta \tau_{j-1}} f(R(\tau_{j-1}), u_{j-1})\} \\ &+ e^{-\beta T} f(R(T), \mu(T)) - e^{-\beta \tau_{N(T)}} f(X(\tau_{N(T)}), u_{N(T)}) \\ &+ \sum_{j=1}^{N(T)} \{e^{-\beta \tau_j} f(R(\tau_j), u_j) - e^{-\beta \tau_j} f(R(\tau_j-), u_{j-1})\}. \end{aligned} \quad (15)$$

Since R is a continuous martingale in the interval $[\tau_{j-1}, \tau_j)$ and $f(\cdot, u_{j-1})$ is $C^2([0, \infty) \setminus \mathcal{N})$, where \mathcal{N} is a finite subset of $[0, \infty)$, then we may apply an appropriate version of Itô formula. Thus, for every $i \in \mathbb{N}$,

$$\begin{aligned} & e^{-\beta \tau_j} f(R(\tau_j-), u_{j-1}) - e^{-\beta \tau_{j-1}} f(R(\tau_{j-1}), u_{j-1}) \\ &= \int_{[\tau_{j-1}, \tau_j)} e^{-\beta t} \mathcal{L} f(R(t), u_{j-1}) dt + \int_{[\tau_{j-1}, \tau_j)} e^{-\beta t} \sigma f'(R(t), u_{j-1}) dB(t) \\ &+ \int_{[\tau_{j-1}, \tau_j)} e^{-\beta t} f'(R(t), u_{j-1}) dA(t) \\ &= \int_{[\tau_{j-1}, \tau_j)} e^{-\beta t} \mathcal{L} f(R(t), u_{j-1}) dt + \int_{[\tau_{j-1}, \tau_j)} e^{-\beta t} \sigma f'(R(t), u_{j-1}) dB(t) \\ &+ \int_{[\tau_{j-1}, \tau_j)} e^{-\beta t} f'(0, u_{j-1}) dA(t). \end{aligned}$$

The last equality holds because $A(t)$ increases only when $R(t) = 0$. According to

inequality (12) and the fact that $f(\cdot, \mu_i)$ is a solution for the QVI in (7), we have

$$\begin{aligned} & \mathbb{E}_{x,i} [e^{-\beta\tau_j} f(R(\tau_j-), u_{j-1}) - e^{-\beta\tau_{j-1}} f(R(\tau_{j-1}), u_{j-1})] \\ & \geq \mathbb{E}_{x,i} \left[\int_{[\tau_{j-1}, \tau_j)} e^{-\beta t} (-c(u_{j-1}) - hR(t)) dt \right] + \mathbb{E}_{x,i} \left[\int_{[\tau_{j-1}, \tau_j)} e^{-\beta t} (-U) dA(t) \right], \end{aligned} \quad (16)$$

and

$$\mathbb{E}_{x,i} [e^{-\beta\tau_j} (f(R(\tau_j), u_j) - f(R(\tau_j-), u_{j-1}))] \geq \mathbb{E}_{x,i} [e^{-\beta\tau_j} (-K(u_{j-1}, u_j))]. \quad (17)$$

We note that the above two inequalities become equalities for the QVI-control associated with $\{f(\cdot, \mu_i), i = 1, \dots, k\}$, and $f'(0, \mu_i) = -U$ if $0 \in \mathcal{C}_i, i = 1, 2, \dots, k$.

Combining the above inequalities, and applying them to equation (15), we obtain

$$\begin{aligned} & \mathbb{E}_{x,i} [e^{-\beta T} f(R(T), \mu(T))] - f(x, \mu_i) \\ & \geq \mathbb{E}_{x,i} \left[\int_0^T e^{-\beta t} (-c(\mu(t)) - hR(t)) dt \right] + \mathbb{E}_{x,i} \left[\int_0^T e^{-\beta t} (-U) dA(t) \right] \\ & \quad + \sum_{j=1}^{N(T)} \mathbb{E}_{x,i} [e^{-\beta\tau_j} (-K(u_{j-1}, u_j))]. \end{aligned}$$

The boundedness of f' implies

$$f(x, \mu_i) \leq M(1 + |x|), \quad i = 1, 2, \dots, k,$$

which further implies

$$f(R(T), \mu(T)) \leq M(1 + |R(T)|) \leq M(1 + |x| + \max_{i=1, \dots, k} |\mu_i| \cdot T + \sigma |B(T)| + A(T)). \quad (18)$$

Since (4) implies that

$$\liminf_{T \rightarrow \infty} \mathbb{E}[e^{-\beta T} A(T)] = 0,$$

then we obtain

$$\begin{aligned} & \liminf_{T \rightarrow \infty} (\mathbb{E}_{x,i} [e^{-\beta T} f(R(T), \mu(T))] - f(x, \mu_i)) \\ & \geq \liminf_{T \rightarrow \infty} \left(\mathbb{E}_{x,i} \left[\int_0^T e^{-\beta t} (-c(\mu(t)) - hR(t)) dt \right] + \mathbb{E}_{x,i} \left[\int_0^T e^{-\beta t} (-U) dA(t) \right] \right) \\ & \quad + \liminf_{T \rightarrow \infty} \left(\sum_{j=1}^{N(T)} \mathbb{E}_{x,i} [e^{-\beta\tau_j} (-K(u_{j-1}, u_j))] \right), \end{aligned}$$

which implies

$$f(x, \mu_i) \leq DC(x, \mu_i, \Phi),$$

with equality when Φ is the QVI-control associated with $\{f(\cdot, \mu_i), i = 1, \dots, k\}$, and $f'(0, \mu_i) = -U$ if $0 \in \mathcal{C}_i, i = 1, 2, \dots, k$. \square

2.5 *Explicit solution in the two drift case*

In this section, we consider the case of two drifts, $k = 2$, for a one dimensional state process in $[0, \infty)$ as in the previous section. We first assume that $\mu_1 \geq 0, \mu_2 < 0$. We restrict the feasible policies to be drift controls as given by (5). A drift control band policy is defined by two parameters a, b with $0 < a < b$. Under the policy, if the current drift is μ_1 , the reserve authority changes the drift to μ_2 when the reserve level rises to b ; if the current drift is μ_2 , the reserves authority changes the drift to μ_1 when the reserve level falls to a . Given a control band policy Φ^D , we provide a method for policy evaluation. We then establish the existence of a solution to the free boundary problem with parameters (a^*, b^*) under certain conditions. Finally, under certain conditions, the control band policy associated with (a^*, b^*) is indeed optimal among all feasible policies.

2.5.1 Control band policies

Let $\{a, b\}$ denote the control band policy associated with parameters a, b with $0 < a < b$. Fix a policy $\Phi^D = \{a, b\}$, the adjustment time and adjustment quantity are

given by the following: For initial state $(x, u_0) = (x, \mu_1)$, we have

$$\begin{aligned}
\tau_1 &= \inf\{t \geq \tau_0 : R(t) \geq b\}, \\
u_1 &= \mu_2, \\
\tau_2 &= \inf\{t > \tau_1 : R(t) \leq a\}, \\
u_1 &= \mu_1, \\
&\vdots \\
n &\geq 1 \\
\tau_{2n+1} &= \inf\{t > \tau_{2n} : R(t) \geq b\}, \\
u_{2n+1} &= \mu_2, \\
\tau_{2n+2} &= \inf\{t > \tau_{2n+1} : R(t) \leq a\}, \\
u_{2n+1} &= \mu_1,
\end{aligned} \tag{19}$$

Or for initial state $(x, u_0) = (x, \mu_2)$, we have

$$\begin{aligned}
\tau_1 &= \inf\{t \geq \tau_0 : R(t) \leq a\}, \\
u_1 &= \mu_1, \\
\tau_2 &= \inf\{t > \tau_1 : R(t) \geq b\}, \\
u_1 &= \mu_2, \\
&\vdots \\
n &\geq 1 \\
\tau_{2n+1} &= \inf\{t > \tau_{2n} : R(t) \leq a\}, \\
u_{2n+1} &= \mu_1, \\
\tau_{2n+2} &= \inf\{t > \tau_{2n+1} : R(t) \geq b\}, \\
u_{2n+1} &= \mu_2,
\end{aligned} \tag{20}$$

and the continuation and intervention regions are explicitly characterized by

$$\mathcal{C}_1 = [0, b), \quad \mathcal{S}_{12} = [b, \infty).$$

$$\mathcal{C}_2 = (a, \infty), \quad \mathcal{S}_{21} = [0, a].$$

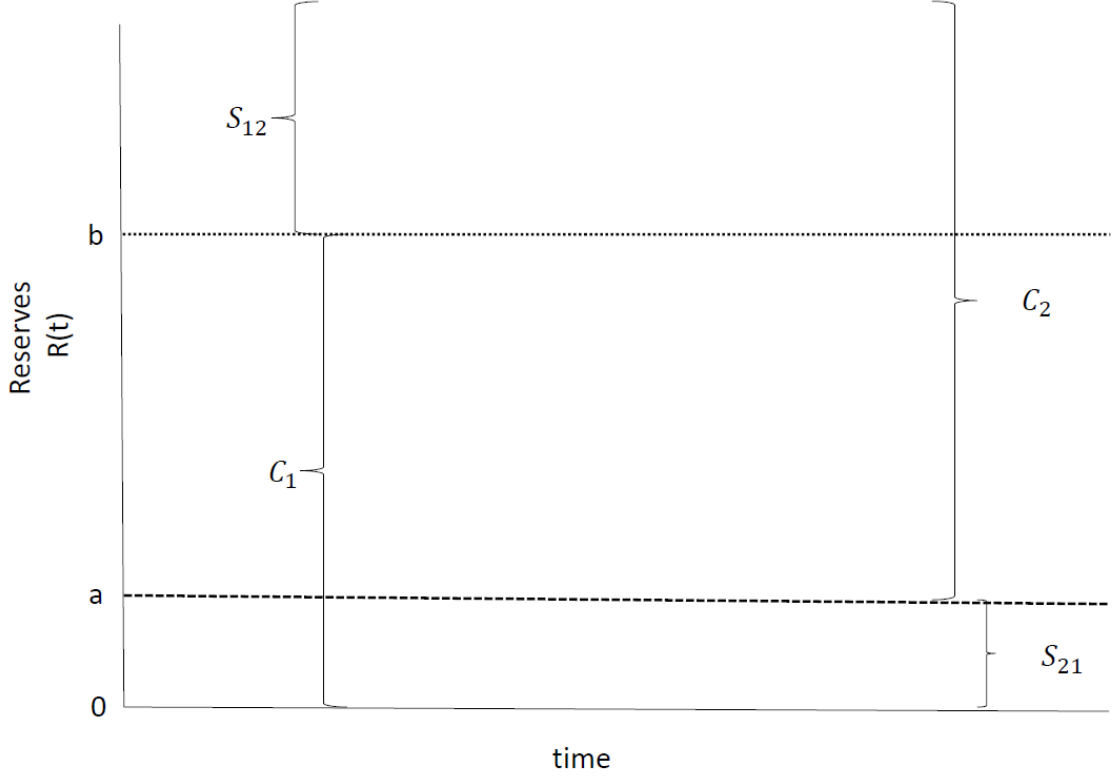


Figure 2: The continuation and intervention regions in the two drift case

In the following theorem, we obtain an expression for the *value function* $DC(x, \mu_i, \Phi^D)$, the expected total discounted cost resulted from a control band policy when the initial reserve level and initial drift are x and μ_i , respectively.

Theorem 2.5.1. *Assume that we fix a control band policy $\Phi^D = \{a, b\}$. If there exists twice continuously differentiable functions $w_1 : [0, b) \rightarrow \mathbb{R}$ and $w_2 : (a, \infty) \rightarrow \mathbb{R}$*

that satisfy

$$\mathcal{L}_1 w_1(x) + hx + c(\mu_1) = 0, \quad x \in [0, b) \quad (21)$$

$$\mathcal{L}_2 w_2(x) + hx + c(\mu_w) = 0, \quad x \in (a, \infty) \quad (22)$$

$$w'_2(x) \text{ is bounded in } (a, \infty), \quad (23)$$

with boundary conditions

$$w_1(b) = w_2(b) + K(\mu_1, \mu_2), \quad (24)$$

$$w_2(a) = w_1(a) + K(\mu_2, \mu_1), \quad (25)$$

$$w'_1(b) = w'_2(b), \quad (26)$$

$$w'_2(a) = w'_1(a), \quad (27)$$

$$w'_1(0) = -U. \quad (28)$$

then the expected total discounted cost is given by

$$DC(x, \mu_1, \Phi^D) = \begin{cases} w_1(x) & \text{for } x \in [0, b), \\ w_2(x) + K(\mu_1, \mu_2) & \text{for } x \in [b, \infty), \end{cases}$$

and

$$DC(x, \mu_2, \Phi^D) = \begin{cases} w_1(x) + K(\mu_2, \mu_1) & \text{for } x \in [0, a], \\ w_2(x) & \text{for } x \in (a, \infty), \end{cases}$$

where $\{w_i(x), i = 1, 2\}$ are defined in (21)-(22).

Proof. Let's define

$$f(x, \mu_1) = \begin{cases} w_1(x) & \text{for } x \in [0, b), \\ w_2(x) + K(\mu_1, \mu_2) & \text{for } x \in [b, \infty), \end{cases}$$

and

$$f(x, \mu_2) = \begin{cases} w_1(x) + K(\mu_2, \mu_1) & \text{for } x \in [0, a], \\ w_2(x) & \text{for } x \in (a, \infty), \end{cases}$$

then equations (21)-(22) and (24)-(27) imply that

$$\begin{aligned} f(\cdot, \mu_1) &\in C^1([0, \infty); [0, \infty)) \cap C^2([0, \infty) \setminus \{b\}; [0, \infty)), \\ f(\cdot, \mu_2) &\in C^1([0, \infty); [0, \infty)) \cap C^2([0, \infty) \setminus \{a\}; [0, \infty)), \end{aligned}$$

and they are the solution of the QVI. Meanwhile, equations (23) and (28) imply that $\{f(x, \mu_i), i = 1, 2\}$ are bounded and $f'(0, \mu_1) = -U$, respectively.

Similar to the proof in Theorem 2.4.1, we have

$$DC(x, \mu_i, \Phi^D) = f(x, \mu_i), \quad i = 1, 2.$$

□

Remark 1. *Equations (24)-(28) guarantee that value functions $\{C(x, \mu_i, \Phi^D), i = 1, 2\}$ are C^1 on $[0, \infty)$. These are also known as the “smooth-pasting” condition.*

To facilitate the presentation, we explicitly find a solution to (21)-(22). Define

$$\begin{aligned} m_i^+ &= \frac{1}{2} \left(-\frac{2\mu_i}{\sigma^2} + \sqrt{\left(\frac{2\mu_i}{\sigma^2}\right)^2 + 4\frac{2\beta}{\sigma^2}} \right) > 0, \quad i = 1, 2 \\ m_i^- &= \frac{1}{2} \left(-\frac{2\mu_i}{\sigma^2} - \sqrt{\left(\frac{2\mu_i}{\sigma^2}\right)^2 + 4\frac{2\beta}{\sigma^2}} \right) < 0, \quad i = 1, 2. \end{aligned}$$

For $i = 1, 2$, the general solution to the ordinary differential equation

$$\mathcal{L}_i w_i(x) + hx + c(\mu_i) = 0, \tag{29}$$

is

$$w_i(x) = A_i \varphi_i(x) + B_i \psi_i(x) + \phi_i(x),$$

where A_i, B_i are real numbers,

$$\varphi_i(x) = e^{m_i^+ x}, \quad \psi_i(x) = e^{m_i^- x},$$

and

$$\phi_i(x) = \frac{h}{\beta} x + \frac{h\mu_i}{\beta^2} + \frac{c(\mu_i)}{\beta}.$$

We end this section by providing a criterion, which is equivalent to the smooth-pasting condition (24)-(28), and easy to verify numerically.

Corollary 2.5.2. *For a given control band policy $\Phi^D = \{a, b\}$, suppose the following equation holds:*

$$\begin{aligned} & \begin{pmatrix} \psi_1(a) - \frac{m_1^-}{m_1^+} \varphi_1(a) & -\psi_2(a) \\ \psi_1'(a) - \frac{m_1^-}{m_1^+} \varphi_1'(a) & -\psi_2'(a) \end{pmatrix}^{-1} \begin{pmatrix} \phi_2(a) - \phi_1(a) + K_{12} + \frac{\frac{h}{\beta} + U}{m_1^+} \varphi_1(a) \\ \phi_2'(a) - \phi_1'(a) + \frac{\frac{h}{\beta} + U}{m_1^+} \varphi_1'(a) \end{pmatrix} \\ & - \begin{pmatrix} \psi_1(b) - \frac{m_1^-}{m_1^+} \varphi_1(b) & -\psi_2(b) \\ \psi_1'(b) - \frac{m_1^-}{m_1^+} \varphi_1'(b) & -\psi_2'(b) \end{pmatrix}^{-1} \begin{pmatrix} \phi_2(b) - \phi_1(b) + K_{12} + \frac{\frac{h}{\beta} + U}{m_1^+} \varphi_1(b) \\ \phi_2'(b) - \phi_1'(b) + \frac{\frac{h}{\beta} + U}{m_1^+} \varphi_1'(b) \end{pmatrix} = 0. \end{aligned} \quad (30)$$

Then the following functions

$$H_1(x) = \begin{cases} A_1 \varphi_1(x) + B_1 \psi_1(x) + \phi_1(x) & \text{for } x \in [0, b], \\ B_2 \psi_2(x) + \phi_2(x) + K(\mu_1, \mu_2) & \text{for } x \in [b, \infty), \end{cases}$$

and

$$H_2(x) = \begin{cases} A_1 \varphi_1(x) + B_1 \psi_1(x) + \phi_1(x) + K(\mu_2, \mu_1) & \text{for } x \in [0, a], \\ B_2 \psi_2(x) + \phi_2(x) & \text{for } x \in (a, \infty), \end{cases}$$

are the expected total discounted cost corresponding to the given control band policy

$\Phi^D = \{a, b\}$, where the coefficients A_1, B_1, B_2 are given by

$$\begin{aligned} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} &= \begin{pmatrix} \psi_1(a) - \frac{m_1^-}{m_1^+} \varphi_1(a) & -\psi_2(a) \\ \psi_1'(a) - \frac{m_1^-}{m_1^+} \varphi_1'(a) & -\psi_2'(a) \end{pmatrix}^{-1} \begin{pmatrix} \phi_2(a) - \phi_1(a) + K_{12} + \frac{\frac{h}{\beta} + U}{m_1^+} \varphi_1(a) \\ \phi_2'(a) - \phi_1'(a) + \frac{\frac{h}{\beta} + U}{m_1^+} \varphi_1'(a) \end{pmatrix} \\ &= \begin{pmatrix} \psi_1(b) - \frac{m_1^-}{m_1^+} \varphi_1(b) & -\psi_2(b) \\ \psi_1'(b) - \frac{m_1^-}{m_1^+} \varphi_1'(b) & -\psi_2'(b) \end{pmatrix}^{-1} \begin{pmatrix} \phi_2(b) - \phi_1(b) + K_{12} + \frac{\frac{h}{\beta} + U}{m_1^+} \varphi_1(b) \\ \phi_2'(b) - \phi_1'(b) + \frac{\frac{h}{\beta} + U}{m_1^+} \varphi_1'(b) \end{pmatrix}. \end{aligned}$$

and

$$A_1 = \frac{-U - \phi_1'(0) - B_1 \psi_1'(0)}{\varphi_1(0)}.$$

Moreover, $\{H_i(x), i = 1, 2\}$ are in $C^1([0, \infty); [0, \infty)) \cap C^2([0, \infty) \setminus \mathcal{N}; [0, \infty))$ where \mathcal{N} is a finite subset of $[0, \infty)$, and $\{H_i'(x), i = 1, 2\}$ are bounded in $[0, \infty)$.

Proof. By Theorem 2.5.1, if functions $\{H_i(x), i = 1, 2\}$ satisfy conditions (21)-(28), then they are the expected total discounted cost corresponding to the given control band policy $\Phi^D = \{a, b\}$.

Conditions (21)-(22) hold automatically by definition. Furthermore, condition $m_2^- < 0$ implies that

$$B_2\psi'_2(x) + \phi'_2(x) = B_2m_2^-e^{m_2^-x} + \frac{h}{\beta}$$

is bounded in (a, ∞) .

Combine with equation (30), if we define

$$\begin{aligned} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} &= \begin{pmatrix} \psi_1(a) - \frac{m_1^-}{m_1^+}\varphi_1(a) & -\psi_2(a) \\ \psi'_1(a) - \frac{m_1^-}{m_1^+}\varphi'_1(a) & -\psi'_2(a) \end{pmatrix}^{-1} \begin{pmatrix} \phi_2(a) - \phi_1(a) + K_{12} + \frac{\frac{h}{\beta}+U}{m_1^+}\varphi_1(a) \\ \phi'_2(a) - \phi'_1(a) + \frac{\frac{h}{\beta}+U}{m_1^+}\varphi'_1(a) \end{pmatrix} \\ &= \begin{pmatrix} \psi_1(b) - \frac{m_1^-}{m_1^+}\varphi_1(b) & -\psi_2(b) \\ \psi'_1(b) - \frac{m_1^-}{m_1^+}\varphi'_1(b) & -\psi'_2(b) \end{pmatrix}^{-1} \begin{pmatrix} \phi_2(b) - \phi_1(b) + K_{12} + \frac{\frac{h}{\beta}+U}{m_1^+}\varphi_1(b) \\ \phi'_2(b) - \phi'_1(b) + \frac{\frac{h}{\beta}+U}{m_1^+}\varphi'_1(b) \end{pmatrix}, \end{aligned}$$

and

$$A_1 = \frac{-U - \phi'_1(0) - B_1\psi'_1(0)}{\varphi_1(0)},$$

then the boundary conditions

$$A_1\varphi_1(b) + B_1\psi_1(b) + \phi_1(b) = B_2\psi_2(b) + \phi_2(b) + K(\mu_1, \mu_2),$$

$$B_2\psi_2(a) + \phi_2(a) = A_1\varphi_1(a) + B_1\psi_1(a) + \phi_1(a) + K(\mu_2, \mu_1),$$

$$A_1\varphi'_1(b) + B_1\psi'_1(b) + \phi'_1(b) = B_2\psi'_2(b) + \phi'_2(b),$$

$$B_2\psi'_2(a) + \phi'_2(a) = A_1\varphi'_1(a) + B_1\psi'_1(a) + \phi'_1(a),$$

$$A_1\varphi'_1(0) + B_1\psi'_1(0) + \phi'_1(0) = -U,$$

hold automatically. Furthermore, these boundary conditions imply that $\{H_i(x), i = 1, 2\}$ are in $C^1([0, \infty); [0, \infty)) \cap C^2([0, \infty) \setminus \mathcal{N}; [0, \infty))$ where \mathcal{N} is a finite subset of $[0, \infty)$. \square

Remark 2. The coefficient B_2 can be represented as

$$\begin{aligned}
B_2 &= \frac{1}{\det(a)} [m_1^-(e^{m_1^+a} - e^{m_1^-a}) (\frac{h(\mu_2 - \mu_1)}{\beta^2} + \frac{c(\mu_2) - c(\mu_1)}{\beta} - K_{21}) \\
&\quad + \frac{m_1^+ - m_1^-}{m_1^+} (\frac{h}{\beta} + U) e^{(m_1^+ + m_1^-)a}] \\
&= \frac{1}{\det(b)} [m_1^-(e^{m_1^+b} - e^{m_1^-b}) (\frac{h(\mu_2 - \mu_1)}{\beta^2} + \frac{c(\mu_2) - c(\mu_1)}{\beta} + K_{12}) \\
&\quad + \frac{m_1^+ - m_1^-}{m_1^+} (\frac{h}{\beta} + U) e^{(m_1^+ + m_1^-)b}],
\end{aligned} \tag{31}$$

where

$$\det(x) = \begin{vmatrix} \psi_1(x) - \frac{m_1^-}{m_1^+} \varphi_1(x) & -\psi_2(x) \\ \psi_1'(x) - \frac{m_1^-}{m_1^+} \varphi_1'(x) & -\psi_2'(x) \end{vmatrix}, \quad \text{for } x \in (0, \infty). \tag{32}$$

2.6 Optimal policy and optimal parameters

Theorem 2.4.1 suggests the following strategy for obtaining the optimal policy. We conjecture that the optimal policy is of the control band form policy. Therefore, the first task is to find the optimal policy among all control band policies. We denote this optimal control band policy by $\Phi^* = \{a^*, b^*\}$ with the expected total discounted cost

$$H_1(x) := DC(x, \mu_1, \Phi^*) = \begin{cases} A_1 \varphi_1(x) + B_1 \psi_1(x) + \phi_1(x) & \text{for } x \in [0, b^*), \\ H_2(x) + K(\mu_1, \mu_2) & \text{for } x \in [b^*, \infty), \end{cases} \tag{33}$$

and

$$H_2(x) := DC(x, \mu_2, \Phi^*) = \begin{cases} H_1(x) + K(\mu_2, \mu_1) & \text{for } x \in [0, a^*], \\ B_2 \psi_2(x) + \phi_2(x) & \text{for } x \in (a^*, \infty). \end{cases} \tag{34}$$

We hope that $\{H_i(x), i = 1, 2\}$ can be used as the functions $\{f(x, \mu_i), i = 1, 2\}$ in Theorem 2.4.1. To find the corresponding $\{f(x, \mu_i), i = 1, 2\}$ that satisfy all of the conditions in Theorem 2.4.1, we derive the conditions that should be imposed on the optimal parameters.

Corollary 2.5.2 provides a criterion to check whether functions $\{H_i(x), i = 1, 2\}$ satisfy the smooth-pasting condition. We then prove in the following Theorem that if $B_2 > 0$, the functions $\{H_i(x), i = 1, 2\}$ are a solution of the QVI, therefore, Φ^* is optimal among all feasible policies.

Theorem 2.6.1. *Let $0 < a^* < b^*$ be the solution of equation (30) in corollary 2.5.2. Furthermore, if $B_2 > 0$, which is defined in (31), then the functions $\{H_i(x), i = 1, 2\}$ defined in (33)-(34) are the optimal value functions in equation (6), and the corresponding control band policy $\Phi^* = \{a^*, b^*\}$ is optimal among all non-anticipated policies.*

Proof. If $0 \leq a^* < b^*$ is the solution of equation (30), then by corollary 2.5.2, the functions $\{H_i(x), i = 1, 2\}$ defined in equations (33) -(34) are in $C^1([0, \infty); [0, \infty)) \cap C^2([0, \infty) \setminus \mathcal{N}; [0, \infty))$ where \mathcal{N} is a finite subset of $[0, \infty)$. Moreover, $\{H'_i(x), i = 1, 2\}$ are bounded in $[0, \infty)$ with $H'_1(0) = -U$.

To prove Φ^D is the QVI-control associated with $\{H_i(x), i = 1, 2\}$, we need to show

$$\begin{aligned} \mathcal{L}_1 H_1(x) + c(\mu_1) + hx &= 0, & \text{for } x \in [0, b^*), \\ H_2(x) + K(\mu_1, \mu_2) &> H_1(x), & \text{for } x \in [0, b^*), \end{aligned} \quad (35)$$

and

$$\begin{aligned} \mathcal{L}_2 H_2(x) + c(\mu_2) + hx &= 0, & \text{for } x \in (a^*, \infty), \\ H_1(x) + K(\mu_2, \mu_1) &> H_2(x), & \text{for } x \in (a^*, \infty). \end{aligned} \quad (36)$$

According to the definition of $\{H_i(x), i = 1, 2\}$, the above conditions are equivalent to the following inequalities

$$\begin{aligned} g_{12}(x) &= H_1(x) - H_2(x) - K(\mu_1, \mu_2) < 0, & \text{for } x \in (a^*, b^*), \\ g_{21}(x) &= H_2(x) - H_1(x) - K(\mu_2, \mu_1) < 0, & \text{for } x \in (a^*, b^*). \end{aligned}$$

Since

$$g_{12}(a^*) = H_1(a^*) - H_2(a^*) - K(\mu_1, \mu_2) = -K(\mu_2, \mu_1) - K(\mu_1, \mu_2) < 0,$$

$$g_{12}(b^*) = H_1(b^*) - H_2(b^*) - K(\mu_1, \mu_2) = 0,$$

and

$$g_{21}(a^*) = H_2(a^*) - H_1(a^*) - K(\mu_2, \mu_1) = 0,$$

$$g_{21}(b^*) = H_2(b^*) - H_1(b^*) - K(\mu_2, \mu_1) = -K(\mu_1, \mu_2) - K(\mu_2, \mu_1) < 0,$$

then the inequality (35) (or, (36)) holds automatically if we can show $g_{12}(x)$ (or, $g_{21}(x)$) is monotonically increasing (or, decreasing) in (a^*, b^*) , which is equivalent to $g'_{12}(x) > 0$ (or, $g'_{21}(x) < 0$) for all $x \in (a^*, b^*)$. Meanwhile,

$$g'_{12}(x) = -g'_{21}(x), \text{ for } x \in (a^*, b^*)$$

implies that we only need to show

$$g'_{12}(x) < 0, \text{ for } x \in (a^*, b^*).$$

As $\{H_i(x), i = 1, 2\}$ are C^1 , we get

$$H'_1(a^*) = H'_2(a^*), \quad H'_1(b^*) = H'_2(b^*)$$

which implies

$$g'_{21}(a^*) = -g'_{21}(b^*) = 0.$$

Therefore, $g'_{21}(x) < 0$ for all $x \in (a^*, b^*)$, as required, provided that $g'_{21}(x)$ can not attain nonnegative maximum in (a, b) . This is indeed the case if we show

$$\mathcal{L}_1 g'_{21}(x) > 0, \quad \text{for } x \in (a^*, b^*), \tag{37}$$

and then appeal to the maximum principle. Simple calculation shows that

$$\mathcal{L}_1 g'_{21}(x) = B_2 m_2^2 (\mu_1 - \mu_2) e^{m_2^- x}.$$

Therefore, the condition $B_2 > 0$ implies that the inequality (37) holds. Now functions $\{H_i(x), i = 1, 2\}$ satisfy all of the conditions in Theorem 2.4.1. Hence, $\{H_i(x), i = 1, 2\}$ are the optimal value functions in equation (6), and the corresponding control band policy $\Phi^* = \{a^*, b^*\}$ is optimal among all non-anticipated policies. \square

In order to apply Theorem 2.6.1, one needs to verify if the corresponding algebraic equation (30) admits solutions, and $B_2 > 0$. The former one is easy to check numerically, as shown in the next section. The remainder of this section will focus on checking the condition of $B_2 > 0$.

Lemma 2.6.2. *For any $x > 0$, $\det(x) > 0$, where $\det(x)$ is defined in (32).*

Proof. According to equation (32), $\det(x) > 0$ is equivalent to

$$e^{(m_1^+ - m_1^-)x} > \frac{m_1^+(m_1^- - m_2^-)}{m_1^-(m_1^+ - m_2^-)}.$$

$$1. \ m_1^- > m_2^-.$$

We have

$$e^{(m_1^+ - m_1^-)x} > 0 > \frac{m_1^+(m_1^- - m_2^-)}{m_1^-(m_1^+ - m_2^-)}.$$

$$2. \ m_1^- > m_2^-.$$

Since

$$m_1^+ \cdot m_1^- - m_2^- \cdot m_1^- < m_1^+ \cdot m_1^- - m_2^- \cdot m_1^+ \leq 0,$$

then we have

$$e^{(m_1^+ - m_1^-)x} > 1 > \frac{m_1^+(m_1^- - m_2^-)}{m_1^-(m_1^+ - m_2^-)}.$$

Therefore, we conclude that $\det(x) > 0$ for $x > 0$. \square

Corollary 2.6.3. *Let $0 < a^* < b^*$ be the solution of equation (30) in corollary 2.5.2 which satisfies*

$$\begin{aligned} & m_1^-(e^{m_1^+ a^*} - e^{m_1^- a^*}) \left(\frac{h(\mu_2 - \mu_1)}{\beta^2} + \frac{c(\mu_2) - c(\mu_1)}{\beta} - K_{21} \right) \\ & + \frac{m_1^+ - m_1^-}{m_1^+} \left(\frac{h}{\beta} + U \right) e^{(m_1^+ + m_1^-)a^*} > 0, \end{aligned} \quad (38)$$

or

$$m_1^-(e^{m_1^+b^*} - e^{m_1^-b^*}) \left(\frac{h(\mu_2 - \mu_1)}{\beta^2} + \frac{c(\mu_2) - c(\mu_1)}{\beta} + K_{12} \right) + \frac{m_1^+ - m_1^-}{m_1^+} \left(\frac{h}{\beta} + U \right) e^{(m_1^+ + m_1^-)b^*} > 0, \quad (39)$$

then the functions $\{H_i(x), i = 1, 2\}$ defined in (33)-(34) are the optimal value functions in equation (6), and the corresponding control band policy $\Phi^* = \{a^*, b^*\}$ is optimal among all non-anticipated policies.

Proof. The proof is just the application of Theorem 2.6.1 and Lemma 2.6.2. \square

Remark 3. In order to apply Corollary 2.6.3, one needs to evaluate if the corresponding algebraic equations (30) admit solutions. This is easy to verify numerically, as shown in Section 2.7.

2.7 Numerical Results

In this section, we provide numerical examples illustrating the optimality of the drift control policy as well as sensitivity analysis. In the first part, we validate the optimality of the drift control policy by comparing its cost function with that of a suboptimal policy. In the second part, we do the sensitivity analysis of the optimal costs and the threshold levels with respect to different model parameters.

2.7.1 Optimal Cost Comparison

As mentioned in [2], a martingale approach was exploited to derive a closed-form expression for the cost function for the threshold type policies. Specifically, the model in [2] assume that at time 0 the reserve level is a and the drift is γ_0 . The drift is switched to γ_1 the first time reserves hit level b . The drift is controlled back to γ_0 as soon as reserves hit level a again, and so forth. The model also assumes that the cost of holding \$1 of reserves per unit of time is h , and the cost per \$1 of regulation

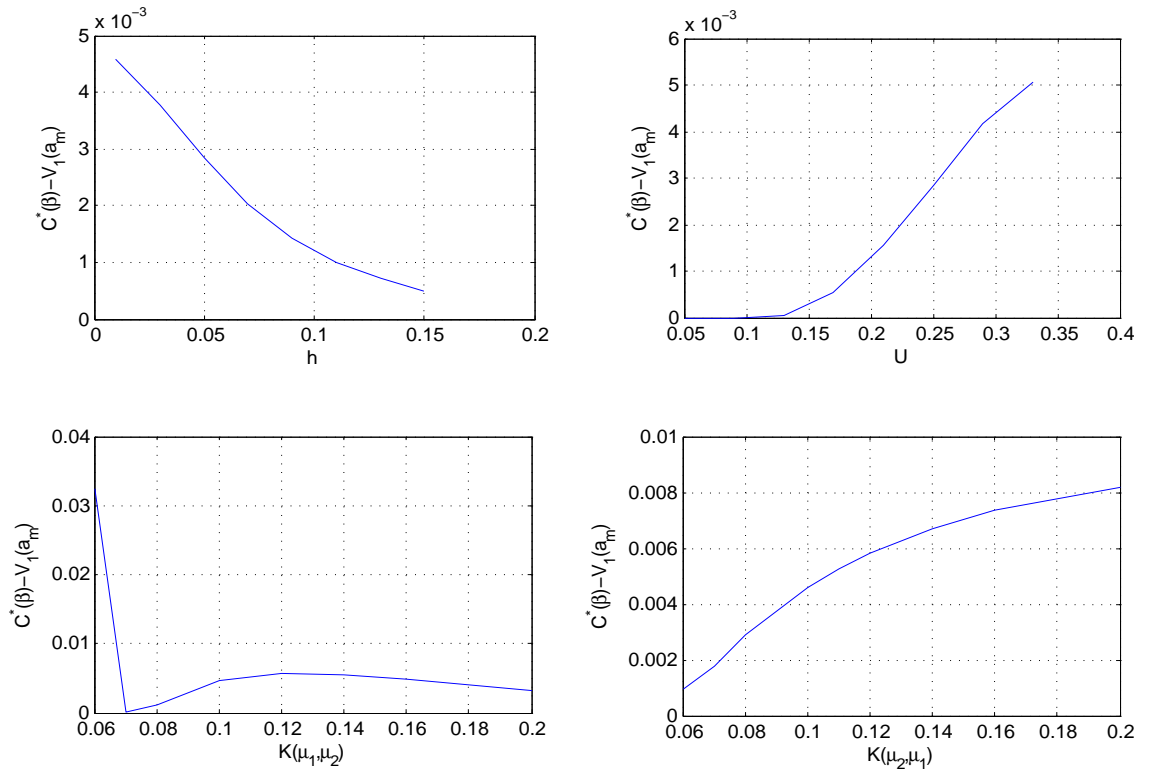


Figure 3: The difference between $c^*(\beta)$ and $V_1(a_m)$ for different parameters $h, U, K(\mu_1, \mu_2)$, and $K(\mu_2, \mu_1)$ in two drift case

at the boundary 0 is k . A cost π_1 is incurred every time the drift is switched from γ_0 to γ_1 , and a cost π_0 is incurred when the drift is switched from γ_1 to γ_0 .

Notations $(k, \pi_0, \pi_1, \gamma_0, \gamma_1)$ in [2] correspond to notations $(U, K(\mu_2, \mu_1), K(\mu_1, \mu_2), \mu_1, \mu_2)$ in our problem formulation.

For each set of parameters $h, k, \pi_0, \pi_1, \sigma$, [2] obtains an analytic solution for the cost function

$$C(\beta, a, b, \gamma_0, \gamma_1) = h \frac{\mathbb{E}_a \left[\int_0^{T_0} e^{-\beta t} R(t) dt \right] + \theta_0(\beta) \mathbb{E}_b \left[\int_0^{T_0} e^{-\beta t} R(t) dt \right]}{1 - \theta_0(\beta) \theta_1(\beta)} \\ + k \frac{\mathbb{E}_a \left[\int_0^{T_0} e^{-\beta t} dL(t) \right]}{1 - \theta_0(\beta) \theta_1(\beta)} + \frac{\pi_1 \theta_0(\beta) + \pi_0 \theta_0(\beta) \theta_1(\beta)}{1 - \theta_0(\beta) \theta_1(\beta)},$$

where $\theta_0(\beta) = \mathbb{E}_a [e^{-\beta T_0}]$, $\theta_1(\beta) = \mathbb{E}_b [e^{-\beta T_1}]$, and $\mathbb{E}_a[\cdot] = \mathbb{E}[\cdot | R(0) = z]$. The vector of triggers and drifts $(a_m, b_m, \gamma_0^*, \gamma_1^*)$ that minimizes the total discounted cost

$$C^*(\beta) = \arg \min_{(a, b, \gamma_0, \gamma_1)} C(\beta, a, b, \gamma_0, \gamma_1) = C(\beta, a_m, b_m, \gamma_0^*, \gamma_1^*)$$

are numerically computed.

We then substitute

$$(\gamma_0^*, \gamma_1^*, h, k, \pi_0, \pi_1, \sigma),$$

which correspond to

$$(\mu_1^*, \mu_2^*, h, U, K(\mu_2, \mu_1), K(\mu_1, \mu_2), \sigma)$$

in our problem, into our optimal control problem with $c(\mu_1) = c(\mu_2) = 0$, and we obtain the optimal value function

$$V_i(x) = \arg \min_{\phi \in \Phi} DC(x, \mu_i^*, \phi) \leq DC(x, \mu_i^*, \phi), \quad \forall \phi \in \Phi, \quad i = 1, 2.$$

with the optimal control band policy $\{a^*, b^*\}$.

From the construction of policy space in [2], it is expected that their optimal policy is suboptimal for our problem due to the fact that the policy space in their problem is a constrained subset of the policy space in our problem: it contains all

the control band policies with lower threshold being equal to the starting point of the reserves. Therefore, by setting the starting point and initial drift of reserves as a_m and μ_1^* , we have

$$V_1(a_m) \leq C^*(\beta).$$

The following set of parameters is chosen as our baseline.

$$(h, U, c_1, c_2, c_3, K(\mu_1, \mu_2), K(\mu_1, \mu_3), K(\mu_2, \mu_3), K(\mu_2, \mu_1), K(\mu_3, \mu_2), K(\mu_3, \mu_1), \beta, \sigma) \\ = (0.07, 0.4, 0, 0, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.04, 1).$$

We compare our closed-form solution with the martingale approach solution, when one of the model parameters varies. The difference between $V_1(a_m)$ and $C^*(\beta)$ in Figure 3 validates the conclusion that the drift control policy is optimal among all of the non-anticipated policies.

2.7.2 Sensitivity Analysis

Here we do the sensitivity analysis of the optimal costs (with initial drift μ_1 and starting point $x = 0.2$) and the threshold levels with respect to $h, U, K(\mu_1, \mu_2)$, and $K(\mu_2, \mu_1)$, when one of the parameters varies. Here, the drift rates are fixed as

$$(\mu_1, \mu_2) = (0.515916, -1.25482).$$

We first vary h and keep all other parameters fixed. The resulting optimal cost $V_1(0.2)$ and the threshold levels (a^*, b^*) are listed in Table 1. The optimal cost increases, while both threshold levels a^* and b^* decrease with the increase in h . This shows that a larger h leads to a higher expected cost, and therefore as lower threshold levels.

We then vary U . The result in Table 2 implies that the optimal cost and the threshold levels increase if U increases. Intuitively, we might expect the reserve authority to react to higher regulation cost by raising the lower threshold level and perhaps raising the upper threshold levels as well.

We finally vary $K(\mu_1, \mu_2)$ and $K(\mu_2, \mu_1)$. Table 3 and Table 4 suggest that as switching costs increase, the lower threshold level a^* decreases while the upper threshold level b^* increases: this is because higher switching drift cost needs to be compensated by a larger gap between the upper and the lower threshold levels.

Table 1: Dependency on h

h	$V_1(0.2)$	a^*	b^*
0.01	1.5189	1.3194	6.3986
0.03	2.7793	0.8570	4.1437
0.05	3.7286	0.6665	3.3842
0.07	4.5379	0.5516	2.9567
0.09	5.2616	0.4717	2.6698
0.11	5.9256	0.4117	2.4588
0.13	6.5450	0.3645	2.2946
0.15	7.1294	0.3260	2.1617
0.20	8.4758	0.2541	1.9147
0.30	10.8463	0.1652	1.6089
0.50	14.8890	0.0735	1.2857

Table 2: Dependency on U

U	$V_1(0.2)$	a^*	b^*
0.20	3.9397	0.2416	2.6467
0.25	4.1084	0.3315	2.7366
0.30	4.2624	0.4119	2.8170
0.35	4.4049	0.4848	2.8899
0.40	4.5379	0.5516	2.9567
0.45	4.6630	0.6132	3.0183
0.50	4.7814	0.6704	3.0755
0.55	4.8941	0.7238	3.1289
0.60	5.0019	0.7740	3.1791

Table 3: Dependency on $K(\mu_1, \mu_2)$

$K(\mu_1, \mu_2)$	$V_1(0.2)$	a^*	b^*
0.03	4.2065	0.6281	2.6903
0.05	4.3069	0.6030	2.7726
0.07	4.4023	0.5808	2.8495
0.10	4.5379	0.5516	2.9567
0.15	4.7476	0.5110	3.1188
0.20	4.9415	0.4776	3.2656
0.25	5.1229	0.4493	3.4011
0.30	5.2942	0.4247	3.5276

2.8 *Explicit solution in the three drifts case*

In this section, we consider the case where the number of drifts is $k = 3$ for a one dimensional state process in $[0, \infty)$ as in the previous section. The drift control

Table 4: Dependency on $K(\mu_2, \mu_1)$

$K(\mu_2, \mu_1)$	$V_1(0.2)$	a^*	b^*
0.03	4.2292	0.6305	2.6934
0.05	4.3230	0.6047	2.7748
0.07	4.4119	0.5818	2.8508
0.10	4.5379	0.5516	2.9567
0.15	4.7317	0.5094	3.1166
0.20	4.9100	0.4745	3.2611
0.25	5.0760	0.4446	3.3943
0.30	5.2321	0.4186	3.5184

problem with only two drifts, the upward and downward drifts of the reserve, is too simple to characterize the real dynamics of the reserves accumulation. Normally, the central bank can adopt more than two types of fiscal policy to change the reserve accumulation: neutral fiscal policy, expansionary fiscal policy, and contractionary fiscal policy. Therefore, we extend the problem to a three drift control case with drifts $\mu_1 \geq 0$, $\mu_2 = 0$, and $\mu_3 < 0$, which is more realistic than the two-drift model for decision making for reserve authorities. As before, the control policies are restricted to drift controls specified in (5). A corresponding drift control band policy is defined by five parameters $y_{12}, y_{13}, x_{21}, y_{23}, x_{31}, x_{32}$ with $0 < x_{21} = x_{31} < x_{32} < y_{12} < y_{13} = y_{23} < \infty$. Under the policy, if the current drift is μ_1 , the reserve authority changes the drift to μ_2 or μ_3 when the reserve level rises to y_{12} or y_{13} ; if the current drift is μ_2 , the reserve authority changes the drift to μ_1 or μ_3 when the reserve level falls to x_{21} or rises to y_{23} ; if the current drift is μ_3 , the reserve authority changes the drift to μ_2 or μ_1 when the reserve level falls to x_{32} or x_{31} . Given a control band policy Φ^D , we provide a method for evaluating the cost of the policy.

2.8.1 Control band policies

Fixing a control band policy $\Phi^D = \{y_{12}, y_{13}, x_{21}, y_{23}, x_{31}, x_{32}\}$ with

$$0 < x_{21} = x_{31} < x_{32} < y_{12} < y_{13} = y_{23} < \infty,$$

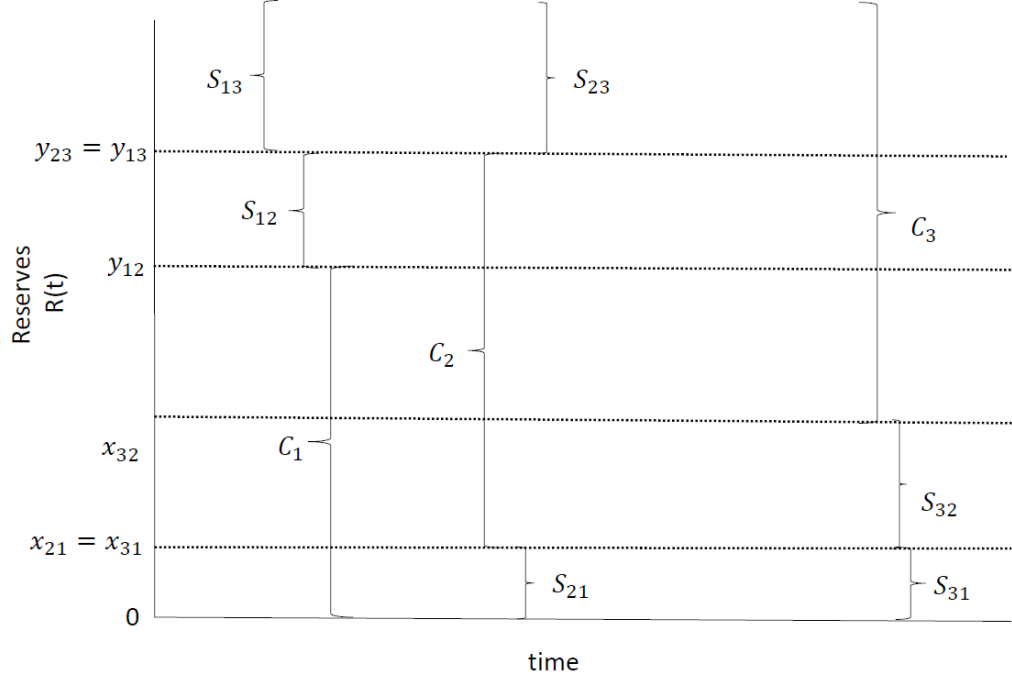


Figure 4: The continuation and intervention regions in the three drift case

the continuation region and the intervention region is explicitly characterized by:

$$\begin{aligned}
\mathcal{C}_1 &= [0, y_{12}), & \mathcal{S}_{12} &= [y_{12}, y_{13}), & \mathcal{S}_{13} &= [y_{13}, \infty). \\
\mathcal{C}_2 &= (x_{21}, y_{23}), & \mathcal{S}_{21} &= [0, x_{21}], & \mathcal{S}_{23} &= [y_{23}, \infty). \\
\mathcal{C}_3 &= (x_{32}, \infty), & \mathcal{S}_{31} &= [0, x_{31}], & \mathcal{S}_{32} &= (x_{31}, x_{32}].
\end{aligned}$$

In the following theorem, we obtain an expression for the *value function* $DC(x, \mu_i, \Phi^D)$, the expected total discounted cost under a control band policy when the initial reserve level and initial drift are x and μ_i , respectively.

Theorem 2.8.1. *Assume that we fix a control band policy $\Phi^D = \{y_{12}, y_{13}, x_{21}, y_{23}, x_{32}, x_{31}\}$ with $0 < x_{21} = x_{31} < x_{32} < y_{12} < y_{13} = y_{23} < \infty$. If there exist twice continuously differentiable functions $w_1 : [0, y_{12}) \rightarrow \mathbb{R}$,*

$w_2 : (x_{21}, y_{23}) \rightarrow \mathbb{R}$, and $w_3 : (x_{32}, \infty) \rightarrow \mathbb{R}$ that satisfy

$$\mathcal{L}_1 w_1(x) + hx + c(\mu_1) = 0, \quad x \in [0, y_{12}) \quad (40)$$

$$\mathcal{L}_2 w_2(x) + hx + c(\mu_2) = 0, \quad x \in (x_{21}, y_{23}) \quad (41)$$

$$\mathcal{L}_2 w_3(x) + hx + c(\mu_3) = 0, \quad x \in (x_{32}, \infty) \quad (42)$$

$$w'_3(x) \text{ is bounded in } (x_{32}, \infty), \quad (43)$$

with boundary conditions

$$w_1(y_{12}) = w_2(y_{12}) + K(\mu_1, \mu_2), \quad (44)$$

$$w_2(x_{21}) = w_1(x_{21}) + K(\mu_2, \mu_1), \quad (45)$$

$$w_1(y_{13}) = w_3(y_{13}) + K(\mu_1, \mu_3), \quad (46)$$

$$w_3(x_{31}) = w_1(x_{31}) + K(\mu_3, \mu_1), \quad (47)$$

$$w_2(y_{23}) = w_3(y_{23}) + K(\mu_2, \mu_3), \quad (48)$$

$$w_3(x_{32}) = w_2(x_{32}) + K(\mu_3, \mu_2), \quad (49)$$

$$w'_1(y_{12}) = w'_2(y_{12}), \quad (50)$$

$$w'_2(x_{21}) = w'_1(x_{21}), \quad (51)$$

$$w'_1(y_{13}) = w'_3(y_{13}), \quad (52)$$

$$w'_3(x_{31}) = w'_1(x_{31}), \quad (53)$$

$$w'_2(y_{23}) = w'_3(y_{23}), \quad (54)$$

$$w'_3(x_{32}) = w'_2(x_{32}), \quad (55)$$

$$w'_1(0) = -U. \quad (56)$$

then the expected total discounted cost is given by

$$DC(x, \mu_1, \Phi^D) = \begin{cases} w_1(x) & \text{for } x \in [0, y_{12}), \\ w_2(x) + K(\mu_1, \mu_2) & \text{for } x \in [y_{12}, y_{13}), \\ w_3(x) + K(\mu_1, \mu_3) & \text{for } x \in [y_{13}, \infty), \end{cases}$$

$$DC(x, \mu_2, \Phi^D) = \begin{cases} w_1(x) + K(\mu_2, \mu_1) & \text{for } x \in [0, x_{21}], \\ w_2(x) & \text{for } x \in (x_{21}, y_{23}), \\ w_3(x) + K(\mu_2, \mu_3) & \text{for } x \in [y_{23}, \infty), \end{cases}$$

and

$$DC(x, \mu_3, \Phi^D) = \begin{cases} w_1(x) + K(\mu_3, \mu_1) & \text{for } x \in [0, x_{31}], \\ w_2(x) + K(\mu_3, \mu_2) & \text{for } x \in (x_{31}, x_{32}], \\ w_3(x) & \text{for } x \in (x_{32}, \infty), \end{cases}$$

where $\{w_i(x), i = 1, 2, 3\}$ are defined in (40)-(42).

Same as in Remark 2 in the previous section, we have

$$w_i(x) = A_i \varphi_i(x) + B_i \psi_i(x) + \phi_i(x), \quad i = 1, 2, 3.$$

where A_i, B_i are real numbers,

$$\varphi_i(x) = e^{m_i^+ x}, \quad \psi_i(x) = e^{m_i^- x}, \quad i = 1, 2, 3.$$

and

$$\phi_i(x) = \frac{h}{\beta} x + \frac{h\mu_i}{\beta^2} + \frac{c(\mu_i)}{\beta}, \quad i = 1, 2, 3.$$

The following corollary explicitly express the smooth-pasting condition as a system of 13 equations and nine unknowns.

Corollary 2.8.2. *For a given control band policy $\Phi^D = \{y_{12}, y_{13}, x_{21}, y_{23}, x_{32}, x_{31}\}$ with $0 < x_{21} = x_{31} < x_{32} < y_{12} < y_{13} = y_{23} < \infty$. If we can find A_1, B_1, A_2, B_2, B_3*

which satisfy the following system of equations

$$A_1\varphi_1(y_{12}) + B_1\psi_1(y_{12}) + \phi_1(y_{12}) = A_2\varphi_2(y_{12}) + B_2\psi_2(y_{12}) + \phi_2(y_{12}) + K(\mu_1, \mu_2), \quad (57)$$

$$A_2\varphi_1(x_{21}) + B_2\psi_2(x_{21}) + \phi_2(x_{21}) = A_1\varphi_1(x_{21}) + B_1\psi_1(x_{21}) + \phi_1(x_{21}) + K(\mu_2, \mu_1), \quad (58)$$

$$A_1\varphi_1(y_{13}) + B_1\psi_1(y_{13}) + \phi_1(y_{13}) = B_3\psi_3(y_{13}) + \phi_3(y_{13}) + K(\mu_1, \mu_3), \quad (59)$$

$$B_3\psi_3(x_{31}) + \phi_3(x_{31}) = A_1\varphi_1(x_{31}) + B_1\psi_1(x_{31}) + \phi_1(x_{31}) + K(\mu_3, \mu_1), \quad (60)$$

$$A_2\varphi_2(y_{23}) + B_2\psi_2(y_{23}) + \phi_2(y_{23}) = B_3\psi_3(y_{23}) + \phi_3(y_{23}) + K(\mu_2, \mu_3), \quad (61)$$

$$B_3\psi_3(x_{32}) + \phi_3(x_{32}) = A_2\varphi_2(x_{32}) + B_2\psi_2(x_{32}) + \phi_2(x_{32}) + K(\mu_3, \mu_2), \quad (62)$$

$$A_1\varphi'_1(y_{12}) + B_1\psi'_1(y_{12}) + \phi'_1(y_{12}) = A_2\varphi'_2(y_{12}) + B_2\psi'_2(y_{12}) + \phi'_2(y_{12}), \quad (63)$$

$$A_2\varphi'_2(x_{21}) + B_2\psi'_2(x_{21}) + \phi'_2(x_{21}) = A_1\varphi'_1(x_{21}) + B_1\psi'_1(x_{21}) + \phi'_1(x_{21}), \quad (64)$$

$$A_1\varphi'_1(y_{13}) + B_1\psi'_1(y_{13}) + \phi'_1(y_{13}) = B_3\psi'_3(y_{13}) + \phi'_3(y_{13}), \quad (65)$$

$$B_3\psi'_3(x_{31}) + \phi'_3(x_{31}) = A_1\varphi'_1(x_{31}) + B_1\psi'_1(x_{31}) + \phi'_1(x_{31}), \quad (66)$$

$$A_2\varphi'_2(y_{23}) + B_2\psi'_2(y_{23}) + \phi'_2(y_{23}) = B_3\psi'_3(y_{23}) + \phi'_3(y_{23}), \quad (67)$$

$$B_3\psi'_3(x_{32}) + \phi'_3(x_{32}) = A_2\varphi'_2(x_{32}) + B_2\psi'_2(x_{32}) + \phi'_2(x_{32}), \quad (68)$$

$$A_1\varphi'_1(0) + B_1\psi'_1(0) + \phi'_1(0) = -U. \quad (69)$$

Then the following functions

$$H_1(x) = \begin{cases} A_1\varphi_1(x) + B_1\psi_1(x) + \phi_1(x) & \text{for } x \in [0, y_{12}), \\ A_2\varphi_2(x) + B_2\psi_2(x) + \phi_2(x) + K(\mu_1, \mu_2) & \text{for } x \in [y_{12}, y_{23}), \\ B_3\psi_3(x) + \phi_3(x) + K(\mu_1, \mu_3) & \text{for } x \in [y_{23}, \infty), \end{cases}$$

$$H_2(x) = \begin{cases} A_1\varphi_1(x) + B_1\psi_1(x) + \phi_1(x) + K(\mu_2, \mu_1) & \text{for } x \in [0, x_{21}], \\ A_2\varphi_2(x) + B_2\psi_2(x) + \phi_2(x) & \text{for } x \in (x_{21}, y_{23}), \\ B_3\psi_3(x) + \phi_3(x) + K(\mu_2, \mu_3) & \text{for } x \in [y_{23}, \infty), \end{cases}$$

and

$$H_3(x) = \begin{cases} A_1\varphi_1(x) + B_1\psi_1(x) + \phi_1(x) + K(\mu_3, \mu_1) & \text{for } x \in [0, x_{31}], \\ A_2\varphi_2(x) + B_2\psi_2(x) + \phi_2(x) + K(\mu_3, \mu_2) & \text{for } x \in (x_{31}, x_{32}], \\ B_3\psi_3(x) + \phi_3(x) & \text{for } x \in [x_{32}, \infty), \end{cases}$$

are the expected total discounted cost corresponding to the given control band policy $\Phi^D = \{y_{12}, y_{13}, x_{21}, y_{23}, x_{32}, x_{31}\}$. Moreover, $\{H_i(x), i = 1, 2, 3\}$ are in $C^1([0, \infty); [0, \infty)) \cap C^2([0, \infty) \setminus \mathcal{N}; [0, \infty))$ where \mathcal{N} is a finite subset of $[0, \infty)$, and $\{H'_i(x), i = 1, 2, 3\}$ are bounded in $[0, \infty)$.

Proof. The proof is similar to the proof of Corollary 2.5.2. □

2.8.2 Optimal policy and optimal parameters

Theorem 2.4.1 suggests the following strategy for obtaining the optimal policy. We conjecture that the optimal policy is of control band form policy. Therefore, we try to find an optimal policy among all control band policies. We denote this optimal control band policy by $\Phi^* = \{y_{12}^*, y_{13}^*, x_{21}^*, y_{23}^*, x_{32}^*, x_{31}^*\}$ with $0 < x_{21}^* = x_{31}^* < x_{32}^* < y_{12}^* < y_{13}^* = y_{23}^* < \infty$, and the corresponding expected total discounted cost

$$H_1(x) := DC(x, \mu_1, \Phi^*) = \begin{cases} A_1\varphi_1(x) + B_1\psi_1(x) + \phi_1(x) & \text{for } x \in [0, y_{12}^*), \\ H_2(x) + K(\mu_1, \mu_2) & \text{for } x \in [y_{12}^*, y_{13}^*), \\ H_3(x) + K(\mu_1, \mu_3) & \text{for } x \in [y_{13}^*, \infty), \end{cases} \quad (70)$$

$$H_2(x) := DC(x, \mu_2, \Phi^*) = \begin{cases} H_1(x) + K(\mu_2, \mu_1) & \text{for } x \in [0, x_{21}^*], \\ A_2\varphi_2(x) + B_2\psi_2(x) + \phi_2(x) & \text{for } x \in (x_{21}^*, y_{23}^*), \\ H_3(x) + K(\mu_2, \mu_3) & \text{for } x \in [y_{23}^*, \infty), \end{cases} \quad (71)$$

and

$$H_3(x) := DC(x, \mu_3, \Phi^*) = \begin{cases} H_1(x) + K(\mu_3, \mu_1) & \text{for } x \in [0, x_{31}^*], \\ H_2(x) + K(\mu_3, \mu_2) & \text{for } x \in (x_{31}^*, x_{32}^*], \\ B_3\psi_3(x) + \phi_3(x) & \text{for } x \in (x_{32}^*, \infty), \end{cases} \quad (72)$$

We hope that $\{H_i(x), i = 1, 2, 3\}$ can be used as the functions $\{f(x, \mu_i), i = 1, 2, 3\}$ in Theorem 2.4.1. To find the corresponding $\{f(x, \mu_i), i = 1, 2, 3\}$ that satisfy all of the conditions in Theorem 2.4.1, we provide the conditions that should be imposed on the optimal parameters.

Corollary 2.8.2 provides a criterion to check whether functions $\{H_i(x), i = 1, 2, 3\}$ satisfy the smooth-pasting condition. We then prove in the following theorem that if $B_3 > 0$, and $A_2(m_2^+)^2 e^{m_2^+ x} + B_2(m_2^-)^2 e^{m_2^- x} > 0, x \in (x_{21}^*, y_{12}^*)$, the functions $\{H_i(x), i = 1, 2, 3\}$ is a solution of the QVI, therefore, Φ^* is optimal among all feasible policies.

Theorem 2.8.3. *Let $0 < x_{21}^* = x_{31}^* < x_{32}^* < y_{12}^* < y_{13}^* = y_{23}^*$ be the solution of equations (57)-(69) in corollary 2.8.2. Furthermore, if*

$$B_3 > 0 \quad \text{and} \quad A_2(m_2^+)^2 e^{m_2^+ x} + B_2(m_2^-)^2 e^{m_2^- x} > 0 \quad \text{for } x \in (x_{21}^*, y_{12}^*). \quad (73)$$

Then the functions $\{H_i(x), i = 1, 2, 3\}$ defined in (70)-(72) are the optimal value functions in equation (6), and the corresponding control band policy $\Phi^ = \{y_{12}^*, y_{13}^*, x_{21}^*, y_{23}^*, x_{32}^*, x_{31}^*\}$ is optimal among all non-anticipated policies.*

Proof. If $0 < x_{21}^* = x_{31}^* < x_{32}^* < y_{12}^* < y_{13}^* = y_{23}^*$ is the solution of equations (57)-(69), then by corollary 2.8.2, the functions $\{H_i(x), i = 1, 2, 3\}$ defined in equations are in $C^1([0, \infty); [0, \infty)) \cap C^2([0, \infty) \setminus \mathcal{N}; [0, \infty))$ where \mathcal{N} is a finite subset of $[0, \infty)$. And $\{H_i'(x), i = 1, 2, 3\}$ are bounded in $[0, \infty)$ with $H_1'(0) = -U$.

To prove Φ^D is the QVI-control associated with $\{H_i(x), i = 1, 2, 3\}$, we need to

show

$$\begin{aligned}
\mathcal{L}_1 H_1(x) + c(\mu_1) + hx &= 0, & \text{for } x \in [0, y_{12}^*) \\
H_1(x) &< \min\{H_2(x) + K(\mu_1, \mu_2), H_3(x) + K(\mu_1, \mu_3)\} & \text{for } x \in [0, y_{12}^*) \\
H_1(x) &= H_2(x) + K(\mu_1, \mu_2) \leq H_3(x) + K(\mu_1, \mu_3), & \text{for } x \in [y_{12}^*, y_{13}^*) \\
H_1(x) &= H_3(x) + K(\mu_1, \mu_3) \leq H_2(x) + K(\mu_1, \mu_2), & \text{for } x \in [y_{13}^*, \infty) \quad (74)
\end{aligned}$$

$$\begin{aligned}
\mathcal{L}_2 H_2(x) + c(\mu_2) + hx &= 0, & \text{for } x \in [x_{21}^*, y_{13}^*) \\
H_2(x) &< \min\{H_1(x) + K(\mu_2, \mu_1), H_3(x) + K(\mu_2, \mu_3)\} & \text{for } x \in [x_{21}^*, y_{13}^*) \\
H_2(x) &= H_1(x) + K(\mu_2, \mu_1) \leq H_3(x) + K(\mu_2, \mu_3), & \text{for } x \in [0, x_{21}^*) \\
H_2(x) &= H_3(x) + K(\mu_2, \mu_3) \leq H_1(x) + K(\mu_2, \mu_1), & \text{for } x \in [y_{13}^*, \infty) \quad (75)
\end{aligned}$$

$$\begin{aligned}
\mathcal{L}_3 H_3(x) + c(\mu_3) + hx &= 0, & \text{for } x \in [x_{32}^*, \infty) \\
H_3(x) &< \min\{H_1(x) + K(\mu_3, \mu_1), H_2(x) + K(\mu_3, \mu_2)\} & \text{for } x \in [x_{32}^*, \infty) \\
H_3(x) &= H_1(x) + K(\mu_3, \mu_1) \leq H_2(x) + K(\mu_3, \mu_2), & \text{for } x \in [0, x_{21}^*) \\
H_3(x) &= H_2(x) + K(\mu_3, \mu_2) \leq H_1(x) + K(\mu_3, \mu_1), & \text{for } x \in (x_{21}^*, x_{32}^*) \quad (76)
\end{aligned}$$

According to the definition of $\{H_i(x), i = 1, 2, 3\}$, the above conditions (74)-(76) are satisfied if we can show

$$\begin{aligned}
w_1(x) &< w_2(x) + K(\mu_1, \mu_2), & \text{for } x \in [x_{21}^*, y_{12}^*) \\
w_2(x) + K(\mu_1, \mu_2) &\leq w_3(x) + K(\mu_1, \mu_3), & \text{for } x \in [y_{12}^*, y_{13}^*) \quad (77)
\end{aligned}$$

$$\begin{aligned}
w_2(x) &< w_1(x) + K(\mu_2, \mu_1), & \text{for } x \in [x_{21}^*, y_{12}^*) \\
w_2(x) &\leq w_3(x) + K(\mu_2, \mu_3), & \text{for } x \in [x_{32}^*, y_{13}^*) \quad (78)
\end{aligned}$$

$$\begin{aligned}
w_3(x) &\leq w_2(x) + K(\mu_3, \mu_2), & \text{for } x \in [x_{32}^*, y_{13}^*] \\
w_2(x) + K(\mu_3, \mu_2) &\leq w_1(x) + K(\mu_3, \mu_1), & \text{for } x \in [x_{21}^*, y_{12}^*]
\end{aligned} \tag{79}$$

In the following, we will show if

$$B_3 > 0 \quad \text{and} \quad A_2(m_2^+)^2 e^{m_2^+ x} + B_2(m_2^-)^2 e^{m_2^- x} > 0 \quad \text{for } x \in (x_{21}^*, y_{12}^*),$$

then (79) holds. The other two can be proved in a similar way.

$$\begin{aligned}
g_{32}(x_{32}^*) &= w_3(x_{32}^*) - w_2(x_{32}^*) - K(\mu_3, \mu_2) = 0, \\
g_{32}(y_{13}^*) &= w_3(y_{13}^*) - w_2(y_{13}^*) - K(\mu_3, \mu_2) = -K(\mu_2, \mu_3) - K(\mu_3, \mu_2) < 0,
\end{aligned}$$

If we can show $g_{32}(x)$ is monotonically decreasing in (x_{32}^*, y_{13}^*) , in other words, $g'_{32}(x) < 0$ for all $x \in (x_{32}^*, y_{13}^*)$, then

$$w_3(x) \leq w_2(x) + K(\mu_3, \mu_2), \quad \text{for } x \in [x_{32}^*, y_{13}^*]$$

holds automatically. By similar argument as that in the proof of Theorem 2.6.1, this is indeed the case if we show

$$\mathcal{L}_2 g'_{32}(x) > 0, \quad \text{for } x \in (x_{32}^*, y_{13}^*),$$

and then appeal to the maximum principle. Simple calculation shows that

$$\mathcal{L}_2 g'_{32}(x) = B_3.$$

Therefore, the condition $B_3 > 0$ implies that

$$w_3(x) \leq w_2(x) + K(\mu_3, \mu_2), \quad \text{for } x \in [x_{32}^*, y_{13}^*].$$

Similarly,

$$\begin{aligned}
g_{21}(x_{31}^*) &= w_2(x_{31}^*) + K(\mu_3, \mu_2) - w_1(x_{31}^*) - K(\mu_3, \mu_1) = 0, \\
g_{21}(y_{12}^*) &= w_2(y_{12}^*) + K(\mu_3, \mu_2) - w_1(y_{12}^*) - K(\mu_3, \mu_1) \\
&= K(\mu_3, \mu_2) - K(\mu_1, \mu_2) - K(\mu_3, \mu_1) < 0,
\end{aligned}$$

We only need to show

$$\mathcal{L}_1 g'_{21}(x) > 0, \quad \text{for } x \in (x_{31}^*, y_{12}^*),$$

which is equivalent to

$$(\mu_1 - \mu_2)[A_2(m_2^+)^2 e^{m_2^+ x} + B_2(m_2^-)^2 e^{m_2^- x}] > 0 \quad \text{for } x \in (x_{31}^*, y_{12}^*).$$

□

Remark 4. *In order to apply Theorem 2.8.3, one needs to evaluate if the corresponding algebraic equations (57)-(69) admit solutions. This is easy to verify numerically, as shown in Section 2.8.3.*

2.8.3 Numerical Results

In this section, we provide numerical examples illustrating the optimality of the drift control policy as well as sensitivity analysis. In the first part, similar to the two-drift case, a suboptimal value function constructed based on martingale approach is used as a benchmark. In the Appendix A.0.4, we derive the closed-form expression (146) for the cost function of the threshold type stopping rules under the three drift assumption. Specifically, for each set of parameters

$$(h, U, K(\mu_1, \mu_2), K(\mu_1, \mu_3), K(\mu_2, \mu_3), K(\mu_2, \mu_1), K(\mu_3, \mu_2), K(\mu_3, \mu_1), \beta, \sigma),$$

we obtain an analytic solution for the cost function in (146). We find numerically the vector of triggers and drifts $(x_{21}^m, y_{12}^m, y_{23}^m, x_{32}^m, \mu_1^*, \mu_2^*, \mu_3^*)$ that minimizes the total discounted cost

$$\begin{aligned} M^*(\beta) &= \arg \min_{(x_{21}, y_{12}, y_{23}, x_{32}, \mu_1, \mu_2, \mu_3)} M(\beta, x_{21}, y_{12}, y_{23}, x_{32}, \mu_1, \mu_2, \mu_3) \\ &= M(\beta, x_{21}^m, y_{12}^m, y_{23}^m, x_{32}^m, \mu_1^*, \mu_2^*, \mu_3^*). \end{aligned}$$

We then substitute

$$(\mu_1^*, \mu_2^*, \mu_3^*, h, U, K(\mu_1, \mu_2), K(\mu_1, \mu_3), K(\mu_2, \mu_3), K(\mu_2, \mu_1), K(\mu_3, \mu_2), K(\mu_3, \mu_1), \beta, \sigma),$$

into our optimal control problem with $c(\mu_1) = c(\mu_2) = c(\mu_3) = 0$, and we obtain the optimal value function

$$V_i(x) = \arg \min_{\phi \in \Phi} DC(x, \mu_i^*, \phi) \leq DC(x, \mu_i^*, \phi), \quad \forall \phi \in \Phi, \quad i = 1, 2, 3.$$

with the optimal control band policy $0 < x_{21}^* = x_{31}^* < x_{32}^* < y_{12}^* < y_{13}^* = y_{23}^* < \infty$.

From the construction of policy space in Appendix A.0.4, it is expected that their optimal policy is suboptimal for our problem due to the fact that the policy space in their problem is a constrained subset of the policy space in our problem: it contains all the control band policies with lower threshold x_{21} being equal to the starting point of the reserves. Therefore, by setting the starting point and initial drift of reserves as x_{21}^m and μ_1^* , we have

$$V_1(x_{21}^m) \leq M^*(\beta).$$

We compare our closed-form solution with the martingale approach solution, when one of the parameters varies. The following set of parameters is chosen as our baseline.

$$(h, U, c_1, c_2, c_3, K(\mu_1, \mu_2), K(\mu_1, \mu_3), K(\mu_2, \mu_3), K(\mu_2, \mu_1), K(\mu_3, \mu_2), K(\mu_3, \mu_1), \beta, \sigma) \\ = (0.07, 0.4, 0, 0, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.04, 1).$$

The difference between $V_1(x_{21}^m)$ and $M^*(\beta)$ in Figure 5 validates the conclusion that the drift control policy is optimal among all of the non-anticipated policies.

In the second part, we do the sensitivity analysis of the optimal costs (with initial drift μ_1 and starting point $x = 0.2$) and the threshold levels with respect to $h, U, K(\mu_1, \mu_2)$, and $K(\mu_2, \mu_3)$, when one of the parameters varies. Here, the drift rates are fixed as

$$(\mu_1, \mu_2, \mu_3) = (0.739756, 0, -1.01397).$$

We first vary h and keep all other parameters fixed. The resulting optimal cost and threshold levels $(y_{12}^*, y_{13}^*, y_{23}^*, x_{21}^*, x_{31}^*, x_{32}^*)$ are listed in Table 5. The optimal cost

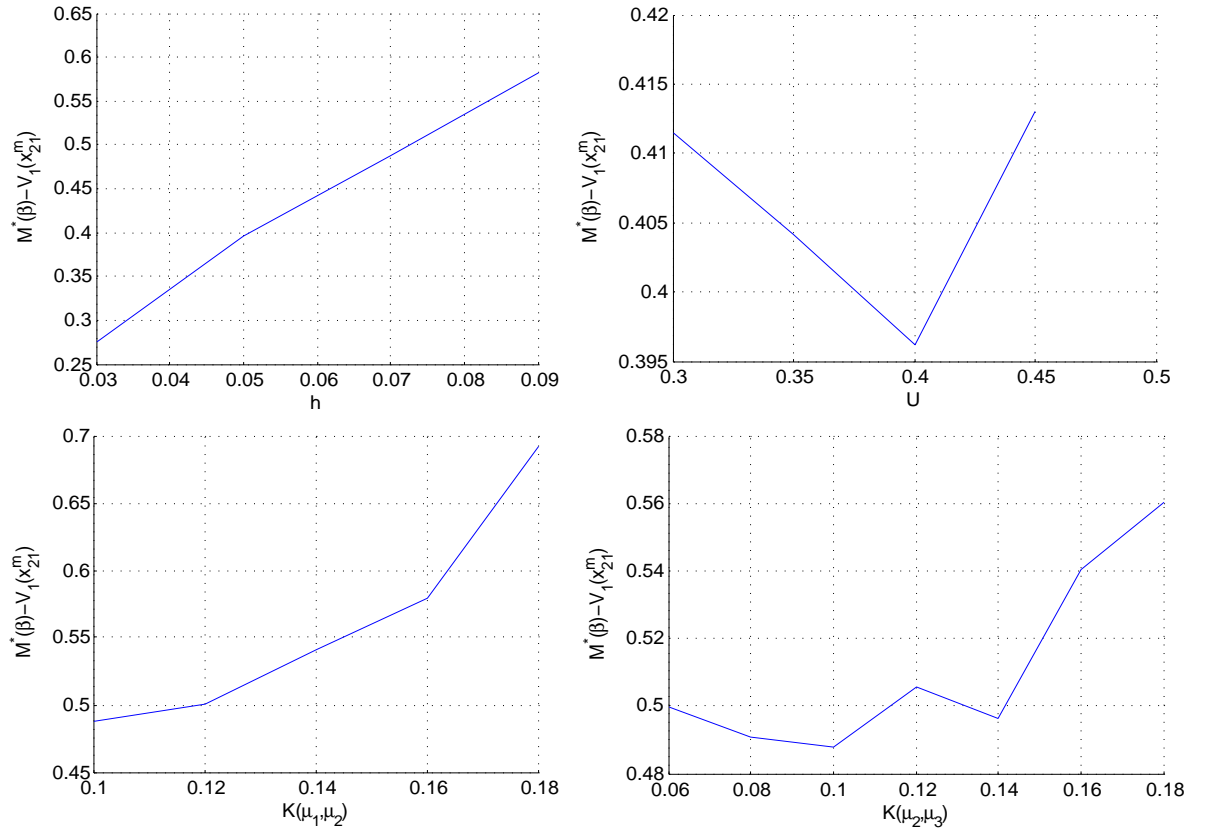


Figure 5: The difference between $M^*(\beta)$ and $V_1(x_{21}^m)$ for different parameters h , U , $K(\mu_1, \mu_2)$, and $K(\mu_2, \mu_3)$ in three drift case

increases while all the threshold levels decrease with the increase in h . This shows that a larger h leads to a higher expected cost, and therefore to lower threshold levels.

We then vary U . The result in Table 6 implies that the optimal cost and all threshold levels increase if U increases. Intuitively, we might expect the reserve authority to react to higher regulation cost by raising the lower threshold levels and perhaps raising the upper threshold levels as well.

We finally vary $K(\mu_1, \mu_2)$ and $K(\mu_2, \mu_3)$. Table 7 and Table 8 suggest that as $K(\mu_1, \mu_2)$ (or, $K(\mu_2, \mu_3)$) increases, the lower threshold level x_{21}^* (or, x_{32}^*) decreases while the upper threshold level y_{12}^* (or, y_{23}^*) increases. This is because higher switching drift cost needs to be compensated for by a larger gap between the upper and the lower threshold levels.

Table 5: Dependency on h

h	y_{12}^*	y_{13}^*	x_{21}^*	y_{23}^*	x_{31}^*	x_{32}^*	$V_1(0.2)$
0.06	2.0894	3.5112	0.1088	3.5112	0.1088	0.9884	4.1436
0.07	1.9874	3.2993	0.0816	3.2993	0.0816	0.9067	4.5645
0.08	1.9024	3.1249	0.0590	3.1249	0.059	0.8393	4.9644
0.09	1.8299	2.9776	0.0398	2.9776	0.0398	0.7823	5.3468
0.10	1.7669	2.8510	0.0233	2.8510	0.0233	0.7332	5.7144
0.11	1.7114	2.7404	0.0088	2.7404	0.0088	0.6903	6.0692

Table 6: Dependency on U

U	y_{12}^*	y_{13}^*	x_{21}^*	y_{23}^*	x_{31}^*	x_{32}^*	$V_1(0.2)$
0.33	1.9099	3.2218	0.0041	3.2218	0.0041	0.8292	4.4124
0.35	1.9331	3.2450	0.0273	3.2450	0.0273	0.8524	4.4575
0.37	1.9554	3.2673	0.0496	3.2673	0.0496	0.8747	4.5645
0.40	1.9874	3.2993	0.0816	3.2993	0.0816	0.9067	4.6252
0.43	2.0178	3.3298	0.1120	3.3298	0.1120	0.9371	4.6252
0.45	2.0373	3.3492	0.1315	3.3492	0.1315	0.9566	4.6644

2.9 Algorithm for an arbitrary number of drifts

In this section, we are going to develop an algorithm to systematically solve the drift control problem with an arbitrary number of drifts. Consider the case where

Table 7: Dependency on $K(\mu_1, \mu_2)$

$K(\mu_1, \mu_2)$	y_{12}^*	y_{13}^*	x_{21}^*	y_{23}^*	x_{31}^*	x_{32}^*	$V_1(0.2)$
0.02	1.6635	3.3041	0.2181	3.3041	0.2181	0.7083	4.3311
0.04	1.7547	3.3096	0.1750	3.3096	0.1750	0.7619	4.3986
0.06	1.8374	3.3102	0.1392	3.3102	0.1392	0.8122	4.4591
0.08	1.9143	3.3066	0.1085	3.3066	0.1085	0.8603	4.5141
0.10	1.9874	3.2993	0.0816	3.2993	0.0816	0.9067	4.5645
0.12	2.0580	3.2886	0.0577	3.2886	0.0577	0.9520	4.6110
0.14	2.1269	3.2744	0.0363	3.2744	0.0363	0.9966	4.6540
0.16	2.1952	3.2570	0.0168	3.2570	0.0168	1.0407	4.6937

Table 8: Dependency on $K(\mu_2, \mu_3)$

$K(\mu_2, \mu_3)$	y_{12}^*	y_{13}^*	x_{21}^*	y_{23}^*	x_{31}^*	x_{32}^*	$V_1(0.2)$
0.02	2.0505	3.1059	0.0919	3.1059	0.0919	0.9547	4.4555
0.04	2.0328	3.1576	0.0891	3.1576	0.0891	0.9414	4.4845
0.06	2.0165	3.2069	0.0865	3.2069	0.0865	0.9290	4.5122
0.08	2.0014	3.2540	0.0840	3.2540	0.0840	0.9175	4.5389
0.10	1.9874	3.2993	0.0816	3.2993	0.0816	0.9067	4.5645
0.12	1.9743	3.3430	0.0794	3.3430	0.0794	0.8966	4.5892
0.14	1.9620	3.3851	0.0773	3.3851	0.0773	0.8871	4.6131
0.16	1.9504	3.4259	0.0753	3.4259	0.0753	0.8781	4.6363
0.18	1.9395	3.4654	0.0733	3.4654	0.0733	0.8695	4.6587

the number of drifts is n for a one-dimensional state process in $[0, \infty)$. We first assume that $\mu_1 > \mu_2 > \dots > \mu_n$, and restrict our feasible policies to non-anticipated policy as in (5). A drift control band policy is defined by parameters $(\{x_{i-h}\}_{h < i, i=1, \dots, n}, \{y_{i-j}\}_{i < j, i=1, \dots, n})$. Under the policy, if the current drift is μ_i , the reserve authority changes the drift to $\mu_h (h < i)$ or $\mu_j (i < j)$ when the reserve level falls to x_{i-h} or rises to y_{i-j} . Given a control band policy, we first provide a method for performance evaluation, and then prove that under certain conditions, the control band policy is optimal among all the non-anticipated policies. Finally, we generalize an algorithm to solve the drift control problem systematically.

2.9.1 Control band policies

Fixing a control band policy $\Phi^D = (\{x_{i\ h}\}_{h < i, i=1, \dots, n}, \{y_{i\ j}\}_{i < j, i=1, \dots, n})$, the continuation region and the intervention region of drift $\mu_i, i = 1, \dots, n$ are explicitly characterized by:

$$\begin{aligned}\mathcal{C}_i &= (x_{i\ i-1}, y_{i\ i+1}), \\ \mathcal{S}_{ij} &= [y_{i\ j}, y_{i\ j+1}), \quad i < j. \\ \mathcal{S}_{ih} &= (x_{i\ h-1}, x_{i\ h}], \quad h < i.\end{aligned}$$

where

$$x_{i\ 0} = 0, \quad y_{i\ n+1} = \infty.$$

The fact that

$$\begin{aligned}S_{ij} &\subset C_j, \quad i < j, \\ S_{ih} &\subset C_h, \quad h < i,\end{aligned}$$

implied by the triangle inequality

$$K(\mu_i, \mu_j) + K(\mu_j, \mu_k) > K(\mu_i, \mu_k)$$

indicates that the boundaries of the continuation and intervention region satisfy the following condition:

$$\begin{aligned}x_{j\ j-1} &< y_{i\ j} < y_{i\ j+1} \leq y_{j\ j+1}, \quad i < j, \\ x_{h\ h-1} &\leq x_{i\ h-1} < x_{i\ h} < y_{h\ h+1}, \quad h < i.\end{aligned}$$

In the following theorem, we obtain an expression for the *value function* $DC(x, \mu_i, \Phi^D)$, the expected total discounted cost under a control band policy when the initial reserve level and initial drift is x and μ_i , respectively.

Theorem 2.9.1. *Assume that we fix a control band policy*

$$\Phi^D = (\{x_{i-h}\}_{h < i, i=1, \dots, n}, \{y_{i-j}\}_{i < j, i=1, \dots, n}).$$

If there exist twice continuously differentiable functions

$$w_i : (x_{i-1}, y_{i+1}) \rightarrow \mathbb{R}, \quad i = 1, 2, \dots, n,$$

that satisfy

$$\mathcal{L}_i w_i(x) + hx + c(\mu_i) = 0, \quad x \in (x_{i-1}, y_{i+1}), \quad i = 1, 2, \dots, n,$$

$$w'_n(x) \text{ is bounded in } (x_{n-1}, \infty),$$

with boundary conditions

$$w_i(x_{i-h}) = w_h(x_{i-h}) + K(\mu_i, \mu_h), \quad i = 2, \dots, n, \quad i > h, \quad (80)$$

$$w'_i(x_{i-h}) = w'_h(x_{i-h}), \quad i = 2, \dots, n, \quad i > h, \quad (81)$$

$$w_i(y_{i-j}) = w_j(y_{i-j}) + K(\mu_i, \mu_j), \quad i = 1, \dots, n-1, \quad i < j, \quad (82)$$

$$w'_i(y_{i-j}) = w'_j(y_{i-j}), \quad i = 1, \dots, n-1, \quad i < j, \quad (83)$$

$$w'_1(0) = -U. \quad (84)$$

then the expected total discounted cost is given by

$$DC(x, \mu_i, \Phi^D) = \begin{cases} w_h(x) + K(\mu_i, \mu_h), & \text{for } x \in (x_{i-h-1}, x_{i-h}], \quad h < i, \\ w_i(x), & \text{for } x \in (x_{i-1}, y_{i+1}), \\ w_j(x) + K(\mu_i, \mu_j), & \text{for } x \in [y_{i-j}, y_{i-j+1}), \quad i < j, \end{cases}$$

where

$$w_i(x) = A_i e^{m_i^+ x} + B_i e^{m_i^- x} + \phi_i(x), \quad i = 1, \dots, n,$$

with $A_n = 0$.

Proof. The proof is similar to the proof of Theorem 2.5.1. □

The above theorem explicitly expresses the smooth-pasting condition as a system of $2n^2 - 2n + 1$ equations with $n^2 + n - 1$ variables. Let's denote the system of equations (80)-(84) by

$$\{f_i(s)\}_{i=1}^{2n^2-2n+1} = 0, \quad i = 1, \dots, 2n(n-1) + 1,$$

where the variable

$$s = (\{x_{ih}\}_{h < i, i=1, \dots, n}, \{y_{ij}\}_{i < j, i=1, \dots, n}, \{A_i\}_{i=1}^{n-1}, \{B_i\}_{i=1}^n).$$

Then

$$f_i(s) = 0, \quad i = 1, \dots, 2n(n-1) + 1,$$

if and only if

$$\sum_{i=1}^{2n(n-1)+1} |f_i(s)|^2 = 0.$$

The following corollary explicitly express the smooth-pasting condition as a least square optimization problem.

Corollary 2.9.2. *For a given control band policy*

$$\Phi^D = (\{x_{ih}\}_{h < i, i=1, \dots, n}, \{y_{ij}\}_{i < j, i=1, \dots, n}).$$

If we can find $\{A_i\}_{i=1}^{n-1}, \{B_i\}_{i=1}^n$ such that the optimal value of the following least-square optimization problem

$$\begin{aligned} \underset{s}{Min} \quad & \sum_{i=1}^{2n(n-1)+1} |f_i(s)|^2 \\ \text{subject to} \quad & x_{jj-1} < y_{ij} < y_{i,j+1} \leq y_{jj+1}, \quad i < j, \end{aligned} \tag{85}$$

$$x_{hh-1} \leq x_{ih-1} < x_{ih} < y_{hh+1}, \quad h < i,$$

is zero, where $\{f_i(s)\}_{i=1}^{2n^2-2n+1}$ are the equations in (80)-(84), and

$$s = (\{x_{ih}\}_{h < i, i=1, \dots, n}, \{y_{ij}\}_{i < j, i=1, \dots, n}, \{A_i\}_{i=1}^{n-1}, \{B_i\}_{i=1}^n),$$

then the following functions

$$H_i(s) = \begin{cases} w_h(x) + K(\mu_i, \mu_h) & \text{for } x \in (x_{i-h-1}, x_{i-h}], \quad h < i, \\ w_i(x) & \text{for } x \in (x_{i-1}, y_{i+1}), \\ w_j(x) + K(\mu_i, \mu_j) & \text{for } x \in [y_{i-j}, y_{i-j+1}), \quad i < j, \end{cases}$$

$i = 1, 2, \dots, n$, are the expected total discounted cost corresponding to the given control band policy

$$\Phi^D = (\{x_{i-h}\}_{h < i, i=1, \dots, n}, \{y_{i-j}\}_{i < j, i=1, \dots, n}),$$

where

$$w_i(x) = A_i e^{m_i^+ x} + B_i e^{m_i^- x} + \phi_i(x), \quad i = 1, \dots, n,$$

with $A_n = 0$, and

$$(\{x_{i-h}\}_{h < i, i=2, \dots, n}, \{y_{i-j}\}_{i < j, i=1, \dots, n-1}, \{A_i\}_{i=1}^{n-1}, \{B_i\}_{i=1}^n)$$

is the optimal solution of the minimization problem (85). Moreover, $\{H_i(x), i = 1, \dots, n\}$ are in $C^1([0, \infty); [0, \infty)) \cap C^2([0, \infty) \setminus \mathcal{N}; [0, \infty))$ where \mathcal{N} is a finite subset of $[0, \infty)$, and $\{H'_i(x), i = 1, \dots, n\}$ are bounded in $[0, \infty)$.

Proof. The proof is similar to the proof of Corollary 2.5.2. □

2.9.2 Optimal policy and optimal parameters

Theorem 2.4.1 suggests the following strategy to obtain an optimal policy. We conjecture that a control band policy is optimal. Therefore, we try to find an optimal policy among all control band policies. We denote this optimal control band policy by

$$\Phi^* = (\{x_{i-h}^*\}_{h < i, i=1, \dots, n}, \{y_{i-j}^*\}_{i < j, i=1, \dots, n}),$$

with

$$x_{j-j-1}^* < y_{i-j}^* < y_{i-j+1}^* \leq y_{j-j+1}^*, \quad i < j,$$

$$x_h^* \leq x_{i-h-1}^* < x_{i-h}^* < y_{h-h+1}^*, \quad h < i,$$

and the corresponding expected total discounted cost

$$H_i(x) := DC(x, \mu_i, \Phi^*) = \begin{cases} w_h(x) + K(\mu_i, \mu_h) & \text{for } x \in (x_{i-h-1}^*, x_{i-h}^*], \quad h < i, \\ w_i(x) & \text{for } x \in (x_{i-1}^*, y_{i+1}^*), \\ w_j(x) + K(\mu_i, \mu_j) & \text{for } x \in [y_{i-j}^*, y_{i-j+1}^*), \quad i < j, \end{cases} \quad (86)$$

for $i = 1, 2, \dots, n$. We hope that $\{H_i(x), i = 1, 2, \dots, n\}$ can be used as the functions $\{f(x, \mu_i), i = 1, 2, \dots, n\}$ in Theorem 2.4.1. To find the corresponding $\{f(x, \mu_i), i = 1, 2, \dots, n\}$ that satisfy all of the conditions in Theorem 2.4.1, Corollary 2.9.2 provides a criterion to check whether functions $\{H_i(x), i = 1, 2, \dots, n\}$ satisfy the smooth-pasting condition. Furthermore, for $i = 1, \dots, n$, if we can prove that

$$H_i(x) < \min\{\{H_h(x) + K(\mu_i, \mu_h)\}_{h < i}, \{H_j(x) + K(\mu_i, \mu_j)\}_{i < j}\}, \quad (87)$$

$$x \in (x_{i-1}^*, y_{i+1}^*),$$

$$H_i(x) = H_h(x) + K(\mu_i, \mu_h) \leq \min\{\{H_g(x) + K(\mu_i, \mu_g)\}_{g < i, g \neq h}, \{H_j(x) + K(\mu_i, \mu_j)\}_{i < j}\}, \quad (88)$$

$$x \in (x_{i-h-1}^*, x_{i-h}^*], \quad h < i,$$

$$H_i(x) = H_j(x) + K(\mu_i, \mu_j) \leq \min\{\{H_h(x) + K(\mu_i, \mu_h)\}_{h < i}, \{H_k(x) + K(\mu_i, \mu_k)\}_{i < k, k \neq j}\}, \quad (89)$$

$$x \in [y_{i-j}^*, y_{i-j+1}^*), \quad i < j,$$

then the functions $\{H_i(x), i = 1, 2, \dots, n\}$ are a solution of the QVI, therefore, Φ^* is optimal among all feasible policies.

In the remainder of this section, we are going to demonstrate how to check the criterion in (87). The other two in (88) and (89) can be done in a similar way.

To check (87) holds, we need to prove

$$H_i(x) < H_h(x) + K(\mu_i, \mu_h), \quad x \in (x_{i-1}^*, y_{i+1}^*) = \mathcal{C}_i, \quad h < i, \quad (90)$$

and

$$H_i(x) < H_j(x) + K(\mu_i, \mu_j), \quad x \in (x_{i-1}^*, y_{i+1}^*) = \mathcal{C}_i, \quad i < j. \quad (91)$$

We only prove (90) in the following sections, (91) follows from the similar arguments.

Case 1: $h = i - 1$

In this case, H_{i-1} can be expressed as

$$H_{i-1}(x) := DC(x, \mu_{i-1}, \Phi^*)$$

$$= \begin{cases} w_g(x) + K(\mu_{i-1}, \mu_g) & \text{for } x \in (x_{i-1-g-1}^*, x_{i-1-g}^*], \quad g < i - 1. \\ w_{i-1}(x) & \text{for } x \in (x_{i-1-i-2}^*, y_{i-1-i}^*), \\ w_j(x) + K(\mu_{i-1}, \mu_j) & \text{for } x \in [y_{i-1-j}^*, y_{i-1-j+1}^*), \quad j > i - 1. \end{cases}$$

with the constraint

$$x_{i-1-i-2}^* < x_{i-i-1}^* < y_{i-1-i}^* < y_{i-1-i+1}^* \leq y_{i-i+1}^*.$$

Let

$$\hat{j} = \max\{j > i - 1 : y_{i-1-j+1}^* \leq y_{i-i+1}^*\},$$

then the inequality

$$H_i(x) < H_{i-1}(x) + K(\mu_i, \mu_{i-1}), \quad x \in (x_{i-i-1}^*, y_{i-i+1}^*),$$

is equivalent to

$$w_i(x) < w_{i-1}(x) + K(\mu_i, \mu_{i-1}), \quad x \in (x_{i-i-1}^*, y_{i-i+1}^*), \quad (92)$$

$$w_i(x) < w_i(x) + K(\mu_{i-1}, \mu_i) + K(\mu_i, \mu_{i-1}), \quad x \in [y_{i-1-i}^*, y_{i-1-i+1}^*), \quad (93)$$

\vdots

$$w_i(x) < w_{\hat{j}}(x) + K(\mu_{i-1}, \mu_{\hat{j}}) + K(\mu_i, \mu_{i-1}), \quad x \in [y_{i-1-\hat{j}}^*, y_{i-1-\hat{j}+1}^*), \quad (94)$$

$$w_i(x) < w_{\hat{j}+1}(x) + K(\mu_{i-1}, \mu_{\hat{j}+1}) + K(\mu_i, \mu_{i-1}), \quad x \in [y_{i-1-\hat{j}+1}^*, y_{i-i+1}^*). \quad (95)$$

Case 2: $h = i - 2$

In this case, H_{i-2} can be expressed as

$$H_{i-2}(x) := DC(x, \mu_{i-2}, \Phi^*)$$

$$= \begin{cases} w_g(x) + K(\mu_{i-2}, \mu_g) & \text{for } x \in (x_{i-2-g-1}^*, x_{i-2-g}^*], \quad g < i - 2, \\ w_{i-2}(x) & \text{for } x \in (x_{i-2-i-3}^*, y_{i-2-i-1}^*), \\ w_j(x) + K(\mu_{i-2}, \mu_j) & \text{for } x \in [y_{i-2-j}^*, y_{i-2-j+1}^*), \quad j > i - 2. \end{cases}$$

- $y_{i-2 \ i-1}^* > x_{i \ i-1}^*$

We have the following constraint

$$x_{i-2 \ i-3}^* < x_{i \ i-1}^* < y_{i-2 \ i-1}^* < y_{i-2 \ i}^* < y_{i-2 \ i+1}^* \leq y_{i \ i+1}^*.$$

Define

$$\hat{j} = \max\{j > i - 2 : y_{i-2 \ j+1}^* \leq y_{i \ i+1}^*\},$$

then the inequality

$$H_i(x) < H_{i-2}(x) + K(\mu_i, \mu_{i-2}), \quad x \in (x_{i \ i-1}^*, y_{i \ i+1}^*),$$

is equivalent to

$$w_i(x) < w_{i-2}(x) + K(\mu_i, \mu_{i-2}), \quad x \in (x_{i \ i-1}^*, y_{i-2 \ i-1}^*), \quad (96)$$

$$w_i(x) < w_{i-1}(x) + K(\mu_{i-2}, \mu_{i-1}) + K(\mu_i, \mu_{i-2}), \quad x \in [y_{i-2 \ i-1}^*, y_{i-2 \ i}^*), \quad (97)$$

\vdots

$$w_i(x) < w_{\hat{j}}(x) + K(\mu_{i-2}, \mu_{\hat{j}}) + K(\mu_i, \mu_{i-2}), \quad x \in [y_{i-2 \ \hat{j}}^*, y_{i-2 \ \hat{j}+1}^*), \quad (98)$$

$$w_i(x) < w_{\hat{j}+1}(x) + K(\mu_{i-2}, \mu_{\hat{j}+1}) + K(\mu_i, \mu_{i-2}), \quad x \in [y_{i-2 \ \hat{j}+1}^*, y_{i \ i+1}^*). \quad (99)$$

- $y_{i-2 \ i-1}^* \leq x_{i \ i-1}^*$

We have the following constraint

$$y_{i-2 \ i-1}^* \leq x_{i \ i-1}^* < y_{i-2 \ i}^* < y_{i-2 \ i+1}^* \leq y_{i \ i+1}^*.$$

Define

$$\hat{j} = \max\{j > i - 2 : y_{i-2 \ j+1}^* \leq y_{i \ i+1}^*\},$$

then the inequality

$$H_i(x) < H_{i-2}(x) + K(\mu_i, \mu_{i-2}), \quad x \in (x_{i-1}^*, y_{i+1}^*)$$

is equivalent to

$$w_i(x) < w_{i-1}(x) + K(\mu_{i-2}, \mu_{i-1}) + K(\mu_i, \mu_{i-2}), \quad x \in (x_{i-1}^*, y_{i-2}^*), \quad (100)$$

$$w_i(x) < w_i(x) + K(\mu_{i-2}, \mu_i) + K(\mu_i, \mu_{i-2}), \quad x \in [y_{i-2}^*, y_{i-2}^*), \quad (101)$$

\vdots

$$w_i(x) < w_{\hat{j}}(x) + K(\mu_{i-2}, \mu_{\hat{j}}) + K(\mu_i, \mu_{i-2}), \quad x \in [y_{i-2}^*, y_{i-2}^*), \quad (102)$$

$$w_i(x) < w_{\hat{j}+1}(x) + K(\mu_{i-2}, \mu_{\hat{j}+1}) + K(\mu_i, \mu_{i-2}), \quad x \in [y_{i-2}^*, y_{i+1}^*). \quad (103)$$

Case 3: $h \leq i - 3$

In this case, H_h can be expressed as

$$H_h(x) := DC(x, \mu_h, \Phi^*) = \begin{cases} w_g(x) + K(\mu_h, \mu_g) & \text{for } x \in (x_{h-g-1}^*, x_{h-g}^*), \quad g < h, \\ w_h(x) & \text{for } x \in (x_{h-1}^*, y_{h+1}^*), \\ w_k(x) + K(\mu_h, \mu_k) & \text{for } x \in [y_{h-k}^*, y_{h-k+1}^*), \quad k > h. \end{cases}$$

- $y_{h+1}^* \leq x_{i-1}^*$

Define

$$j_* = \min\{j > h : y_{h-j+1}^* > x_{i-1}^*\},$$

$$j^* = \max\{j > h : y_{h-j+1}^* \leq y_{i+1}^*\}.$$

Now we have the following constraint

$$y_{h-j_*}^* \leq x_{i-1}^* < y_{h-j_*+1}^* < \cdots < y_{h-j^*+1}^* \leq y_{i+1}^*,$$

and the inequality

$$H_i(x) < H_h(x) + K(\mu_i, \mu_h), \quad x \in (x_{i-1}^*, y_{i+1}^*)$$

is equivalent to

$$w_i(x) < w_{j_*}(x) + K(\mu_h, \mu_{j_*}) + K(\mu_i, \mu_h), \quad x \in (x_{i-1}^*, y_{j_*+1}^*), \quad (104)$$

\vdots

$$w_i(x) < w_{j^*}(x) + K(\mu_h, \mu_{j^*}) + K(\mu_i, \mu_h), \quad x \in [y_{j^*}^*, y_{j^*+1}^*), \quad (105)$$

$$w_i(x) < w_{j^*+1}(x) + K(\mu_h, \mu_{j^*+1}) + K(\mu_i, \mu_h), \quad x \in [y_{j^*+1}^*, y_{i+1}^*). \quad (106)$$

- $y_{h+1}^* > x_{i-1}^*$

Define

$$j^* = \max\{j > h : y_{j+1}^* \leq y_{i+1}^*\}.$$

We have the following constraint

$$x_{h-1}^* \leq x_{i-1}^* < y_{h+1}^* < \cdots < y_{j^*+1}^* \leq y_{i+1}^*,$$

and the inequality

$$H_i(x) < H_h(x) + K(\mu_i, \mu_h), \quad x \in (x_{i-1}^*, y_{i+1}^*)$$

is equivalent to

$$w_i(x) < w_h(x) + K(\mu_i, \mu_h), \quad x \in (x_{i-1}^*, y_{h+1}^*), \quad (107)$$

\vdots

$$w_i(x) < w_{j^*}(x) + K(\mu_h, \mu_{j^*}) + K(\mu_i, \mu_h), \quad x \in [y_{j^*}^*, y_{j^*+1}^*), \quad (108)$$

$$w_i(x) < w_{j^*+1}(x) + K(\mu_h, \mu_{j^*+1}) + K(\mu_i, \mu_h), \quad x \in [y_{j^*+1}^*, y_{i+1}^*). \quad (109)$$

We prove in the following theory that if certain conditions are satisfied, then the inequalities (92)-(95) hold. Similar arguments hold for the criterion in other cases.

Theorem 2.9.3. *If the following conditions hold*

$$(\mu_i - \mu_{i-1})[A_i(m_i^+)^2 e^{m_i^+ x} + B_i(m_i^-)^2 e^{m_i^- x}] > 0, \quad x \in (x_{i-1}^*, y_{i-1}^*), \quad (110)$$

\vdots

$$(\mu_i - \mu_{\hat{j}})[A_i(m_i^+)^2 e^{m_i^+ x} + B_i(m_i^-)^2 e^{m_i^- x}] > 0, \quad x \in (x_{\hat{j}}^*, y_{\hat{j}}^*), \quad (111)$$

$$(\mu_i - \mu_{\hat{j}+1})[A_i(m_i^+)^2 e^{m_i^+ x} + B_i(m_i^-)^2 e^{m_i^- x}] > 0, \quad x \in (x_{\hat{j}+1}^*, y_{\hat{j}+1}^*), \quad (112)$$

then the inequalities (92)-(95) hold automatically.

Proof. Since

$$x_{\hat{j}+1}^* < y_{i-1}^* < y_{i+1}^* \leq y_{\hat{j}+1}^*,$$

and

$$x_{\hat{j}}^* < y_{i-1}^* < y_{i+1}^* < y_{\hat{j}}^*,$$

then (92)-(95) hold if we can show

$$w_i(x) < w_{i-1}(x) + K(\mu_i, \mu_{i-1}), \quad x \in (x_{i-1}^*, y_{i-1}^*), \quad (113)$$

\vdots

$$w_i(x) < w_{\hat{j}}(x) + K(\mu_{i-1}, \mu_{\hat{j}}) + K(\mu_i, \mu_{i-1}), \quad x \in (x_{\hat{j}}^*, y_{\hat{j}}^*), \quad (114)$$

$$w_i(x) < w_{\hat{j}+1}(x) + K(\mu_{i-1}, \mu_{\hat{j}+1}) + K(\mu_i, \mu_{i-1}), \quad x \in (x_{\hat{j}+1}^*, y_{\hat{j}+1}^*). \quad (115)$$

By similar arguments as those in the proof of Theorem 2.8.3, inequalities (113)-(115) hold if conditions (110)-(112) are satisfied. \square

We end this section by describing an algorithm which obtains the optimal control band policy based on the above arguments. The algorithm is used for checking case 1 of criterion (87). Intuitively, step 1 checks the smooth-pasting condition, and step 2 verifies one case of the QVI. Other cases of three different criteria (87), (88) and (89) can be verified similarly.

Algorithm 1 Generic Algorithm for Finding the Optimal Control Band Policy

Inputs: Parameters $(\{\mu_i\}_{i=1}^n, \{c_i\}_{i=1}^n, \{K(\mu_i, \mu_j)\}_{i,j=1, i \neq j}^n, h, U, \beta)$.

Outputs: The optimal control band policy $\Phi^1 = (\{x_{i,h}^1\}_{h < i, i=1, \dots, n}, \{y_{i,j}^1\}_{i < j, i=1, \dots, n})$.

1. Choose an initial point $s^0 = (\{x_{i,h}^0\}_{h < i, i=1, \dots, n}, \{y_{i,j}^0\}_{i < j, i=1, \dots, n}, \{A_i^0\}_{i=1}^{n-1}, \{B_i^0\}_{i=1}^n)$ and an tolerance δ .
 2. Solve the constrained optimization problem (85) given the initial point s^0 and the tolerance δ in step 1. If no solution found, repeat step 1. Else, return the solution $s^1 = (\{x_{i,h}^1\}_{h < i, i=1, \dots, n}, \{y_{i,j}^1\}_{i < j, i=1, \dots, n}, \{A_i^1\}_{i=1}^{n-1}, \{B_i^1\}_{i=1}^n)$, and go to step 3.
 3. If conditions (87)-(89) hold, then $\Phi^1 = (\{x_{i,h}^1\}_{h < i, i=1, \dots, n}, \{y_{i,j}^1\}_{i < j, i=1, \dots, n})$ is the optimal control band policy, return Φ^1 as the output. Otherwise, repeat step 1.
-

CHAPTER III

PARTIAL DISTANCE COVARIANCE: A FAST COMPUTING ALGORITHM

3.1 *Introduction*

Distance covariance and distance correlation are scalar coefficients that characterize independence of random vectors in arbitrary dimension. They were introduced in 2005 by G. J. Székely, in several lectures to address the deficiency of Pearson's correlation, namely that the Pearson's correlation can be zero for dependent variables. We start with a definition of the distance covariance.

Definition 3.1.1. *The population distance covariance between random vectors X and Y with finite first moments is the nonnegative number $\mathcal{V}(X, Y)$ defined by a weighted L_2 norm measuring the distance between the joint characteristic function (c.f.) $\phi_{X,Y}$ of X and Y , and the product $\phi_X\phi_Y$ of the marginal c.f.'s of X and Y . If X and Y take values in \mathbb{R}^p and \mathbb{R}^q , respectively, $\mathcal{V}(X, Y)$ is*

$$\begin{aligned}\mathcal{V}(X, Y) &= \| \phi_{X,Y}(t, s) - \phi_X(t)\phi_Y(s) \|_w^2 \\ &:= \int_{\mathbb{R}^{p+q}} |\phi_{X,Y}(t, s) - \phi_X(t)\phi_Y(s)|^2 w(t, s) dt ds,\end{aligned}$$

where $w(t, s) := (|t|_p^{1+p}|s|_q^{1+q})^{-1}$. The integral exists provided that X and Y have finite first moments.

This shows that distance covariance equals zero if and only if X and Y are independent.

The distance correlation $\mathcal{R}(X, Y)$ is a standardized coefficient, $0 \leq \mathcal{R}(X, Y) \leq 1$,

that also characterizes independence:

$$\mathcal{R}(X, Y) = \begin{cases} \frac{\mathcal{V}(X, Y)}{\sqrt{\mathcal{V}(X, X)\mathcal{V}(Y, Y)}}, & \mathcal{V}(X, X)\mathcal{V}(Y, Y) > 0; \\ 0, & \mathcal{V}(X, X)\mathcal{V}(Y, Y) = 0. \end{cases}$$

The definitions for the corresponding sample statistics have the following simple form.

For an observed random sample $\{(x_i, y_i) : i = 1, \dots, n\}$ from the joint distribution of random vectors X and Y , compute the Euclidean distance matrices $(a_{ij}) = (\|x_i - x_j\|_p)$ and $(b_{ij}) = (\|y_i - y_j\|_q)$, where $\|\cdot\|_p$ denotes the L^p norm in \mathbb{R}^p . Define

$$\hat{A}_{ij} = \begin{cases} a_{ij} - \bar{a}_{i\cdot} - \bar{a}_{\cdot j} + \bar{a}_{\cdot\cdot}, & i, j = 1, \dots, n; \\ 0, & i = j. \end{cases} \quad (116)$$

where

$$\bar{a}_{i\cdot} = \frac{1}{n} \sum_{j=1}^n a_{ij}, \quad \bar{a}_{\cdot j} = \frac{1}{n} \sum_{i=1}^n a_{ij}, \quad \bar{a}_{\cdot\cdot} = \frac{1}{n^2} \sum_{i,j=1}^n a_{ij}.$$

Similarly, define

$$\hat{B}_{ij} = \begin{cases} b_{ij} - \bar{b}_{i\cdot} - \bar{b}_{\cdot j} + \bar{b}_{\cdot\cdot}, & i, j = 1, \dots, n; \\ 0, & i = j. \end{cases} \quad (117)$$

It is clear that the row sums and column sums of these double-centered matrices are 0.

Definition 3.1.2. *The sample distance covariance $\mathcal{V}_n(X, Y)$ and sample distance correlation $\mathcal{R}_n(X, Y)$ are defined by*

$$\mathcal{V}_n^2(X, Y) = \frac{1}{n^2} \sum_{i,j=1}^n \hat{A}_{ij} \hat{B}_{ij} \quad (118)$$

and

$$\tilde{\mathcal{R}}_n^2(X, Y) = \begin{cases} \frac{\tilde{V}_n^2(X, Y)}{\sqrt{\tilde{V}_n^2(X)\tilde{V}_n^2(Y)}}, & \tilde{V}_n^2(X)\tilde{V}_n^2(Y) > 0; \\ 0, & \tilde{V}_n^2(X)\tilde{V}_n^2(Y) = 0. \end{cases} \quad (119)$$

respectively, where the squared sample distance variance is defined by

$$\mathcal{V}_n^2(X) = \mathcal{V}_n^2(X, X) = \frac{1}{n^2} \sum_{i,j=1}^n \hat{A}_{ij}^2. \quad (120)$$

If X and Y have finite first moments, the population distance covariance coefficient $\mathcal{V}^2(X, Y)$ exists and equals zero if and only if the random vectors X and Y are independent. Some of the properties of distance covariance and distance correlation include:

- (i) $\mathcal{V}_n(X, Y)$ and $\mathcal{R}_n(X, Y)$ converge almost surely to $\mathcal{V}(X, Y)$ and $\mathcal{R}(X, Y)$, as $n \rightarrow \infty$.
- (ii) $\mathcal{V}_n(X, Y) \geq 0$ and $\mathcal{V}_n(X) = 0$ if and only if every sample observation is identical.
- (iii) $0 \leq \mathcal{V}_n(X, Y) \leq 1$.
- (iv) If $\mathcal{R}_n(X, Y) = 1$ then there exists a vector a , a nonzero real number b and an orthogonal matrix R such that $Y = a + bXR$, for the data matrices X and Y .

Properties, extensions and applications of distance covariance and distance correlation have been discussed in the recent literature; see, for example, [16] and [18]. A natural question then arises, “How do we define partial distance covariance (correlation) which extends distance covariance (correlation) in a similar sense that partial correlation extends correlation?” One could try to follow the definitions of the classical partial covariance and partial correlation that are based on orthogonal projections in a Euclidean space, but there is a serious difficulty. Orthogonality in case of partial distance covariance and partial distance correlation means independence, but when we compute the orthogonal projection of a random variable onto the condition variable, the “remainder” in the difference is typically not independent of the condition.

Alternately, the form of sample distance covariance (Definition 3.1.2) may suggest an inner product, so one might think of working in the Hilbert space of double centered distance matrices (116), where the inner product is the squared distance covariance statistic (118). Here, we are facing another problem: what would the projections represent? The difference D of double centered distance matrices is typically

not a double centered distance matrix of any sample. This does not affect formal computations, but if we cannot interpret our formulas in terms of samples then inference becomes impossible.

To overcome these difficulties while preserving the essential properties of distance covariance, [27] finally arrived at an elegant solution which starts with defining an alternate type of double centering called \mathcal{U} -centering (see Definition 3.2.1 and Proposition 3.2.2 below). The corresponding inner product is an unbiased estimator of squared population distance covariance. In the Hilbert space of “ \mathcal{U} -centered” matrices, all linear combinations, and in particular projections, are zero diagonal \mathcal{U} -centered matrices.

As a newly developed concept, partial distance covariance has the advantage that it can capture nonlinear dependence [27]. It is expected that partial distance covariance has application in the broad field of life science, engineering, and finance. Particularly as an example, we have successfully applied partial distance covariance to feature screening in high dimensional data; see Chapter 4. If partial distance covariance was implemented directly as it is defined, its computational complexity can be as high as a constant times n^2 for a sample size n . This fact has been regarded as a disadvantage of adopting partial distance covariance. As the main contribution in this part of the thesis, we demonstrate that an $O(n \log n)$ algorithm for a version of the partial distance covariance exists. To illustrate how an $O(n^2)$ order of complexity may hinder the application of an algorithm, assume that n is equal to 10^6 . An $O(n^2)$ algorithm will need 10^{12} numerical operations, which is impossible even for modern computers. In comparison, an $O(n \log n)$ algorithm will only require around 10^6 numerical operations, which is doable. The main idea behind the proposed new algorithm is to use a technique rooted in the the AVL tree structure [1]. The same idea has been utilized to develop a fast algorithm for computing the Kendalls τ rank

correlation coefficient ([15]; [6]), and computing distance covariance ([13]). We extend it to make it suitable for our purpose. The derivation of the fast algorithm also involves significant reformulation from the original version of partial distance covariance. Details are presented in this chapter.

The rest of the chapter is organized as follows: In Section 3.2, we describe the definition of partial distance covariance. Section 3.3 proposes and proves an algorithm with $O(n \log n)$ time complexity. In Section 3.4, we demonstrate the performance of the fast algorithm and compare it with the performance of the $O(n^2)$ algorithm. Finally in Section 3.5 we state the fast algorithm explicitly.

3.2 *Partial distance covariance*

Here we introduce the definition of partial distance covariance statistics. First, we look at “ \mathcal{U} -centered” matrices.

Definition 3.2.1. *Let $A = (a_{ij})$ be a symmetric, real valued $n \times n$ matrix with zero diagonal, $n > 2$. Define the \mathcal{U} -centered matrix \tilde{A} as follows: the (i, j) -th entry of \tilde{A} be*

$$\tilde{A}_{ij} = \begin{cases} a_{ij} - \frac{1}{n-2} \sum_{l=1}^n a_{il} - \frac{1}{n-2} \sum_{k=1}^n a_{kj} + \frac{1}{(n-1)(n-2)} \sum_{k,l=1}^n a_{kl}, & i \neq j; \\ 0, & i = j. \end{cases} \quad (121)$$

“ \mathcal{U} -centered” is so named because as shown below, the corresponding inner product (123) defines an unbiased estimator of squared distance covariance.

Proposition 3.2.2. *Let $(x_i, y_i), i = 1, \dots, n$, denote a sample of observations from the joint distribution (X, Y) of random vectors $X \in \mathcal{R}^p$ and $Y \in \mathcal{R}^q$. Let $A = (a_{ij})$ be the Euclidean distance matrix of the sample x_1, \dots, x_n from the distribution of X , and $B = (b_{ij})$ be the Euclidean distance matrix of the sample y_1, \dots, y_n from the distribution of Y . Then if $E(|X|_p + |Y|_q) < \infty$, for $n > 3$,*

$$(\tilde{A} \cdot \tilde{B}) = \frac{1}{n(n-3)} \sum_{i \neq j} \tilde{A}_{i,j} \tilde{B}_{i,j} \quad (122)$$

is an unbiased estimator of squared population distance covariance $\mathcal{V}^2(X, Y)$.

The proof of the above proposition is in the appendix of [27].

Consider the linear span \mathcal{S}_n of all $n \times n$ distance matrices of samples $\{x_1, \dots, x_n\}$.

Let $A = (a_{ij})$ be an arbitrary element in \mathcal{S}_n . Then A is a real valued, symmetric matrix with zero diagonal. Let $\mathcal{H}_n = \{\tilde{A} : A \in \mathcal{S}_n\}$ and for each pair of elements $C = (C_{i,j}), D = (D_{i,j})$ in the linear span of \mathcal{H}_n define their inner product

$$(C \cdot D) = \frac{1}{n(n-3)} \sum_{i \neq j} C_{ij} D_{ij}. \quad (123)$$

Theorem 3.2.3. *The linear span of all $n \times n$ matrices $\mathcal{H}_n = \{\tilde{A} : A \in \mathcal{S}_n\}$ is a Hilbert space with inner product defined in (123).*

The proof of the above theorem is also in [27].

In what follows, \mathcal{H}_n denotes the Hilbert space of Theorem 3.2.3 with inner product (123), and $|\tilde{A}| = (\tilde{A}, \tilde{A})^{1/2}$ is the norm of \tilde{A} .

3.2.1 Sample pdCov and pdCor

Let \tilde{A}, \tilde{B} , and \tilde{C} be elements of \mathcal{H}_n corresponding to samples x, y and z , respectively, and let

$$P_{z^\perp}(x) = \tilde{A} - \frac{(\tilde{A} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})} \tilde{C}, \quad P_{z^\perp}(y) = \tilde{B} - \frac{(\tilde{B} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})} \tilde{C} \quad (124)$$

denote the orthogonal projection of $\tilde{A}(x)$ onto $(\tilde{C}(z))^\perp$ and the orthogonal projection of $\tilde{B}(y)$ onto $(\tilde{C}(z))^\perp$, respectively. Clearly, $P_{z^\perp}(x)$ and $P_{z^\perp}(y)$ are elements of \mathcal{H}_n , their dot product is defined by (123) and we can define an estimator of $\text{pdCov}(X, Y; Z)$ via projections.

Definition 3.2.4. *(Partial distance covariance) Let (x, y, z) be a random sample observed from the joint distribution of (X, Y, Z) . The sample partial distance covariance (pdCov) is defined by*

$$\text{pdvCov}(x, y; z) = (P_{z^\perp}(x) \cdot P_{z^\perp}(y)), \quad (125)$$

where $P_{z^\perp}(x)$ and $P_{z^\perp}(y)$ are defined by (124), and

$$(P_{z^\perp}(x) \cdot P_{z^\perp}(y)) = \frac{1}{n(n-3)} \sum_{i \neq j} (P_{z^\perp}(x))_{i,j} (P_{z^\perp}(y))_{i,j}. \quad (126)$$

Combined with (124) and (125), we have

$$\begin{aligned} pdvCov(x, y; z) &= ((\tilde{A} - \frac{(\tilde{A} \cdot \tilde{C})}{\tilde{C} \cdot \tilde{C}} \tilde{C}) \cdot (\tilde{B} - \frac{(\tilde{B} \cdot \tilde{C})}{\tilde{C} \cdot \tilde{C}} \tilde{C})) \\ &= (\tilde{A} \cdot \tilde{B}) - \frac{(\tilde{B} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})} (\tilde{A} \cdot \tilde{C}). \end{aligned} \quad (127)$$

If the computation of inner product is implemented directly according to its definition in (123), then the computation complexity of sample partial distance covariance is $O(n^2)$ which is a disadvantage compared to other measures that have lower order of complexity algorithms. Later we will prove that the sample partial distance covariance in (125) can be computed in $O(n \log n)$ steps.

Let Ω_n denote the inner product defined in (122). The following notations will be used. Define the column and row sums as follows:

$$\begin{aligned} a_{i.} &= \sum_{l=1}^n a_{i,l}, & a_{.j} &= \sum_{k=1}^n a_{k,j}, \\ b_{i.} &= \sum_{l=1}^n b_{i,l}, & b_{.j} &= \sum_{k=1}^n b_{k,j}, \\ a_{..} &= \sum_{k,l=1}^n a_{k,l}, & b_{..} &= \sum_{k,j=1}^n b_{k,j}. \end{aligned}$$

Lemma 3.2.5. *If Ω_n is the inner product defined in (122) then we have*

$$\Omega_n = \frac{1}{n(n-3)} \sum_{i \neq j} a_{ij} b_{ij} - \frac{2}{n(n-2)(n-3)} \sum_{i=1}^n a_{i.} b_{i.} + \frac{a_{..} b_{..}}{n(n-1)(n-2)(n-3)}. \quad (128)$$

For the proof see the Appendix. The formula (128) will be used to prove that the estimator in (127) can be computed in $O(n \log n)$ steps.

3.3 An $O(n \log n)$ algorithm

We argue that when X and Y are univariate, there is an $O(n \log n)$ algorithm to implement (128).

Lemma 3.3.1. *Denote*

$$x_{\cdot} = \sum_{i=1}^n x_i.$$

For $1 \leq i \leq n$, we denote

$$\alpha_i^x = \sum_{x_l < x_i} 1,$$

$$\beta_i^x = \sum_{x_l < x_i} x_l.$$

We have

$$a_{i\cdot} = x_{i\cdot} + (2\alpha_i^x - n)x_i - 2\beta_i^x. \quad (129)$$

A proof is relegated to the Appendix

Due to symmetry, the following is the counterpart for Y . We state it without a proof.

Lemma 3.3.2. *Denote*

$$y_{\cdot} = \sum_{i=1}^n y_i.$$

For $1 \leq i \leq n$, we denote

$$\alpha_i^y = \sum_{y_l < y_i} 1,$$

$$\beta_i^y = \sum_{y_l < y_i} y_l.$$

$$b_{i\cdot} = y_{i\cdot} + (2\alpha_i^y - n)y_i - 2\beta_i^y. \quad (130)$$

Using formula (129) and (130), the following two equations can be easily established.

We state them without a proof.

Corollary 3.3.3. *We have*

$$a_{..} = 2 \sum_{i=1}^n \alpha_i^x x_i - 2 \sum_{i=1}^n \beta_i^x, \quad (131)$$

and

$$b_{..} = 2 \sum_{i=1}^n \alpha_i^y y_i - 2 \sum_{i=1}^n \beta_i^y. \quad (132)$$

The following lemma will be used.

Lemma 3.3.4. *We define a sign function, for $\forall 1 \leq i, j \leq n$,*

$$S_{ij} = \begin{cases} +1, & \text{if } (x_i - x_j)(y_i - y_j) > 0, \\ -1, & \text{otherwise.} \end{cases} \quad (133)$$

For any sequence $\{c_j, j = 1, \dots, n\}$, for $1 \leq i \leq n$, we define

$$\gamma_i(\{c_j\}) = \sum_{j:j \neq i} c_j S_{ij}.$$

The following is true:

$$\sum_{i \neq j} a_{ij} b_{ij} = \sum_{i=1}^n [x_i y_i \gamma_i(\{1\}) + \gamma_i(\{x_j y_j\}) - x_i \gamma_i(\{y_j\}) - y_i \gamma_i(\{x_j\})]. \quad (134)$$

Lemma 3.3.5. *For any sequence $\{c_j, j = 1, \dots, n\}$, there is an $O(n \log n)$ algorithm to compute for all $\gamma_i(\{c_j\}) (= \sum_{j:j \neq i} c_j S_{ij})$, where $i = 1, \dots, n$.*

A proof is relegated to the Appendix.

We present the main result in the following theorem.

Theorem 3.3.6. *The sample partial distance covariance that was defined in (125) can be computed by an $O(n \log n)$ algorithm.*

Proof. By (127), the theorem is proved if we can show all the inner products $(\tilde{A} \cdot \tilde{B})$, $(\tilde{A} \cdot \tilde{C})$, $(\tilde{B} \cdot \tilde{C})$, and $(\tilde{C} \cdot \tilde{C})$ can be computed by an $O(n \log n)$ algorithm. We show it for $(\tilde{A} \cdot \tilde{B})$. The other terms can be done in the same way.

In Lemma 3.2.5, the unbiased statistic has been rewritten as in (128). For the first term on the right hand side of (128), per Lemma 3.3.4 and 3.3.5, there is an $O(n \log n)$ algorithm to compute it.

For the second term on the right hand side of (128), note that quantities $\alpha_i^x, \beta_i^x, \alpha_i^y, \beta_i^y$ that were defined in Lemma 3.3.1 and 3.3.2, respectively, are partial sums, which can be computed for all i 's with $O(n \log n)$ algorithms. The $\log n$ factor is inserted, because one may need to sort x_i 's or y_i 's in order to compute for $\alpha_i^x, \beta_i^x, \alpha_i^y, \beta_i^y$. Then by (129) and (130), all a_i and b_i can be computed at order $O(n \log n)$. Consequently, the second term on the right hand side of (123) can be computed by using an $O(n \log n)$ algorithm.

For the third term on the right hand side of (123), using (131) and (132) in Corollary 3.3.3, we can easily see that it can be computed via an $O(n \log n)$ algorithms. From all the above, the theorem is established. \square

3.4 Numerical Results

The fast algorithm was implemented in MATLAB, with the key step of dyadic updating implemented in C. It was then compared against the direct (i.e., slow) implementation. Table 9 presents the average running time for the two different implementations in MATLAB with 2,000 replications at each sample size. The sample size goes from $64 (= 2^6)$ to $8192 (= 2^{13})$. In all these cases, the two methods gave identical solutions, which validates our fast algorithm. Note a comparison in MATLAB is not desirable for our fast algorithm. The direct method calls some MATLAB functions, which achieve the speed of a direct low-level language implementation, while the implementation of the fast method is not. In theory, the fast algorithm will compare more favorably if both methods are implemented in a low-level language, such as in C or C++.

Figure 6 provides a visual comparison of the two methods. All the experiments

that are reported in this chapter are run on a Visual Lab computer (AMD Opteron (tm) Processor 6376 CPU @ 2.30GHz) with allowable 2.72 GB memory in MATLAB Version 8.2.0.89 (R2014a).

When the sample size is large, e.g., when $n = 16,384$, the direct method will generate an “out-of memory” message. Recall that the direct method computes for all pairwise distances, hence it requires $O(n^2)$ memory. The fast method only requires $O(n)$ in memory. For illustration purpose, we run the fast algorithm for sample size n going from 16,384 (which is 2^{14}) to 262,144 (which is 2^{18}). The running times are reported in Table 10. When $n = 262,144$, the running time is a little more than one hundred seconds. The trend that is observable from Figure 7 verifies our claim that the fast method is an $O(n \log n)$ algorithm. It is evident that the running time scales approximately linearly with the sample size (n). We did not run experiments with larger sample sizes, because their outcomes are predictable by the property of the fast method.

Table 9: Running times (in seconds) for the direct and fast methods for computing the partial distance covariances. The values in the parentheses are sample standard errors. At each sample size, 1,000 repetitions were run.

Sample Size	Direct Method	Fast Method
64	0.0015 (0.0092)	0.0076 (0.0062)
128	0.0026 (0.0010)	0.0109 (0.0012)
256	0.0093 (0.0109)	0.0244 (0.0007)
512	0.0527 (0.0106)	0.0567 (0.0022)
1024	0.1929 (0.0163)	0.1339 (0.0065)
2048	0.8007 (0.0506)	0.3248 (0.0376)
4096	3.1398 (0.1177)	0.8130 (0.1037)
8192	12.6775 (0.2447)	1.8926 (0.1781)

3.5 Algorithms

Algorithm 3.5 realizes the idea that is described in the proof of Lemma. Algorithm 3.5 is a subroutine that will be called in Algorithm 3.5. Algorithm 3.5 is the algorithm

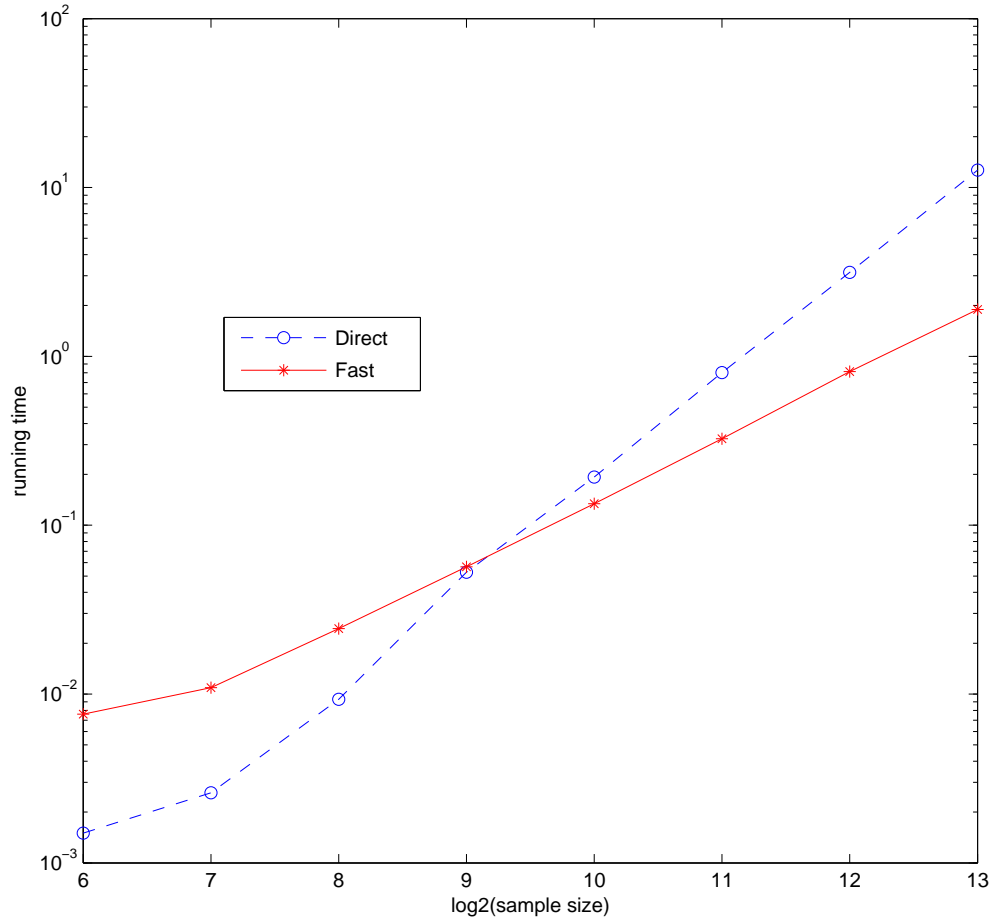


Figure 6: A comparison of running time between the direct and fast methods for the computation of the partial distance covariances

Table 10: Running times (in seconds) for the fast methods for computing the partial distance covariances. The values in the parentheses are sample standard errors. At each sample size, 1,00 repetitions were run.

Sample Size	Fast Method
16384	4.1936 (0.2613)
32768	9.7005 (0.4598)
65536	22.446 (0.6833)
131072	49.1621 (0.9519)
262144	109.1116 (2.5407)

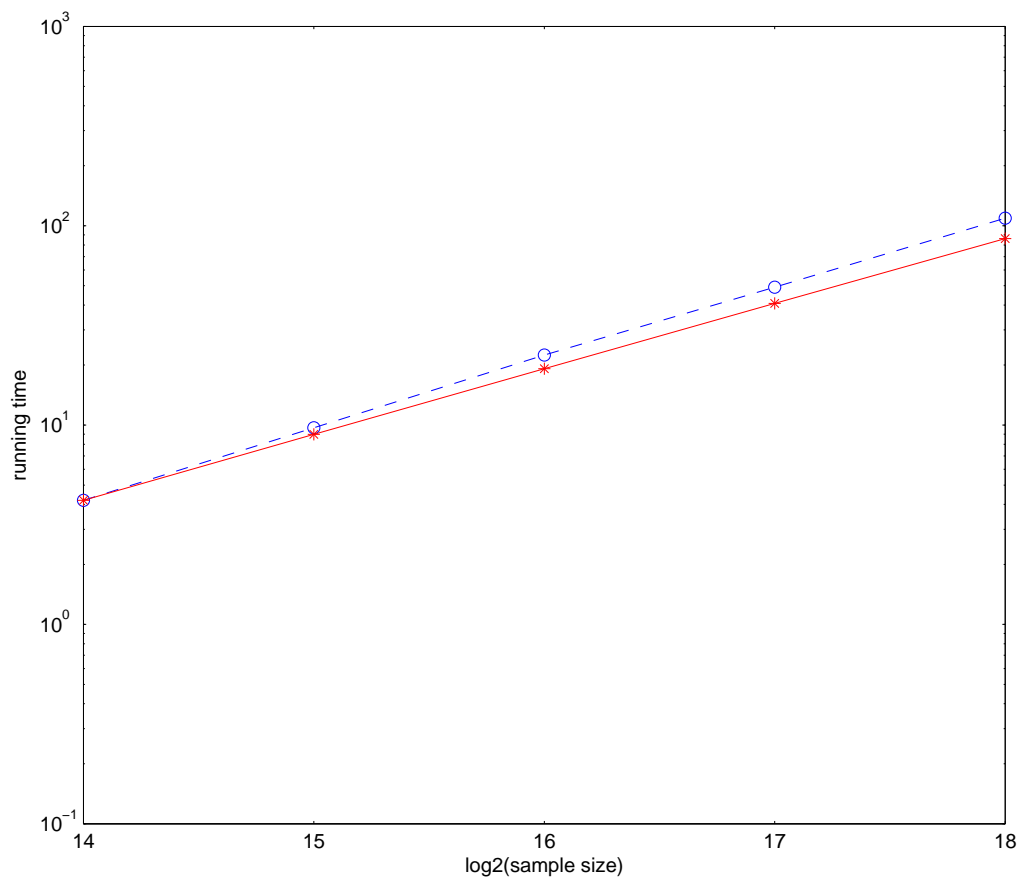


Figure 7: An illustration of running times of the fast method for the computation of the distance correlations. The red solid line corresponds to an $O(n \log n)$ algorithm.

that can compute for the distance covariance at $O(n \log n)$.

Algorithm 2 A subroutine that will be needed in the fast algorithm for the distance covariance. This algorithm realizes the ideas in the proof of Lemma .

Algorithm: Fast Algorithm for a 2-D Partial Sum Sequence (PartialSum2D)

Inputs: Observations x_1, \dots, x_n , y_1, \dots, y_n , and c_1, \dots, c_n .

Outputs: Quantity $\gamma_i(\{c_j\}) = \sum_{j:j \neq i} c_j S_{ij}$ that is defined in Lemma 3.3.4.

1. Compute for the order statistics $x_{(1)} < \dots < x_{(n)}$ for x_1, \dots, x_n . Then rearrange triplets (x_i, y_i, c_j) s such that we have $x_1 < \dots < x_n$. Each triplet (x_i, y_i, c_j) ($1 \leq i \leq n$) stay unchanged.
2. Let $y_{(1)} < \dots < y_{(n)}$ denote the order statistics for y_1, \dots, y_n , and assume that $I^y(i), i = 1, 2, \dots, n$, are the order indices; i.e., if $I^y(i) = k$, then y_i is the k -th smallest among y_1, \dots, y_n . Without loss of generality, we may assume that $y_i = I^y(i)$.
3. Evidently aforementioned function $I^y(i)$ is invertible. Let $(I^y)^{(1)}(j)$ denote its inverse. Define the partial sum sequence: for $1 \leq i \leq n$,

$$s^y(i) = \sum_{j=1}^i c_{(I^y)^{(1)}(j)}.$$

The following recursive relation enables an $O(n)$ algorithm to compute for all $s^y(i)$'s,

$$s^y(1) = c_{(I^y)^{(1)}(1)}, \quad s^y(i+1) = s^y(i) + c_{(I^y)^{(1)}(i+1)}, \quad \text{for } i \geq 1.$$

4. For $1 \leq i \leq n$, define

$$s^x(i) = \sum_{j=1}^i c_j.$$

Again the above partial sums can be computed in $O(n)$ steps.

5. Compute $c. = \sum_{j=1}^n c_j$.
6. Call Subroutine DyadUpdate to compute for $\sum_{j:j < i, y_j < y_i} c_j$ for all $i, 1 \leq i \leq n$.
7. By (151), we have that

$$\gamma_i(\{c_j\}) = c. - c_i - 2s^y(i) - 2s^x(i) + 4 \sum_{j:j < i, y_j < y_i} c_j.$$

Algorithm 3 A subroutine that will be called in Algorithm 3.5.

Subroutine: A Dyadic Updating Scheme (DyadUpdate)

Inputs: Sequence y_1, \dots, y_n and c_1, \dots, c_n , where y_1, \dots, y_n is a permutation of $\{1, \dots, n\}$.

Outputs: Quantities $\gamma_i := \sum_{j:j < i, y_j < y_i} c_j$, $i = 1, 2, \dots, n$.

1. Recall that we have assumed $n = 2L$. If n is not dyadic, we simply choose the smallest L such that $n < 2L$. Recall that for $l = 0, 1, \dots, L - 1$, $k = 1, 2, \dots, 2^{L-l}$, we define a close interval

$$I(l, k) := [(k - 1)2^l + 1, \dots, k2^l].$$

2. Assign $s(l, k) = 0$, $\forall l, k$, and $\gamma_1 = 0$.

3. For $i=2, \dots, n$, we do the following.

- (a) Find all (l, k) 's, such that $y_{i-1} \in I(l, k)$. Then for these (l, k) 's, do update

$$s(l, k) \leftarrow s(l, k) + c_{i-1}$$

- (b) Find nonnegative integers $l_1 > \dots > l_\tau \geq 0$ such that

$$y_i - 1 = 2^{l_1} + \dots + 2^{l_\tau}$$

Let $k_i = 1$. For $j = 2, \dots, \tau$, compute

$$k_j = (2^{l_1} + \dots + 2^{l_{j-1}}) \cdot 2^{-l_j} + 1.$$

- (c) Compute $\gamma_i = \sum_{j=1}^{\tau} s(l_j, k_j)$.
-

Algorithm 4 The $O(n \log n)$ algorithm to compute for the partial distance covariances.

Algorithm: Fast Computing for Distance Covariance (FaDCor)

Inputs: Observations x_1, \dots, x_n , y_1, \dots, y_n , and z_1, \dots, z_n .

Outputs: The partial distance covariance that was defined in (125).

1. Sort x_1, \dots, x_n , and y_1, \dots, y_n . Let I^x and I^y denote the order indices; i.e., if for $i, 1 \leq i \leq n, I^x(i) = k$, then x_i is the k th smallest observations among x_1, \dots, x_n . Similarly if for $i, 1 \leq i \leq n, I^y(i) = k$, then y_i is the k th smallest observations among y_1, \dots, y_n .
2. Let $x_{(1)} < \dots < x_{(n)}$, and $y_{(1)} < \dots < y_{(n)}$ denote the order statistics. Denote the partial sums:

$$s^x(i) = \sum_{j=1}^i x_{(j)}, \quad s^y(i) = \sum_{j=1}^i y_{(j)}, \quad i = 1, \dots, n-1.$$

They can be computed using the following recursive relation: $s^x(1) = x_{(1)}, s^y(1) = y_{(1)}$,

$$s^x(i+1) = s^x(i) + x_{(i+1)}, \quad s^y(i+1) = s^y(i) + y_{(i+1)}, \text{ for } i = 1, \dots, n-1.$$

3. Compute $\alpha_i^x, \alpha_i^y, \beta_i^x$, and β_i^y that are defined in Lemma 3.3.1 and 3.3.2, using the following formula: for $i = 1, \dots, n$, we have

$$\begin{aligned} \alpha_i^x &= I^x(i) - 1, & \alpha_i^y &= I^y(i) - 1, \\ \beta_i^x &= s^x(I^x(i) - 1), & \beta_i^y &= s^y(I^y(i) - 1). \end{aligned}$$

4. Compute x_{\cdot} and y_{\cdot} per their definitions in Lemma 3.3.1 and 3.3.2.
 5. Using (129) and (130), compute $\sum_{i=1}^n a_i b_i$.
 6. Using (131) and (132), compute $a_{\cdot\cdot}$ and $b_{\cdot\cdot}$.
 7. Use Algorithm *PartialSum2D* to compute for $\gamma_i(\{1\}), \gamma_i(\{x_j y_j\}), \gamma_i(\{y_j\})$, and $\gamma_i(\{x_j\})$.
 8. Using (134) to compute $\sum_{i \neq j} a_{ij} b_{ij}$.
 9. Apply the results of steps 6, 7, and 8 to (122).
 10. Repeat the steps of 1 through 9 to compute $(\tilde{A} \cdot \tilde{C}), (\tilde{B} \cdot \tilde{C})$, and $(\tilde{C} \cdot \tilde{C})$.
 11. Using formula (127) to compute (125).
-

CHAPTER IV

APPLICATION OF PARTIAL DISTANCE COVARIANCE IN HIGH DIMENSIONAL FEATURE SCREENING

4.1 *Introduction*

Statisticians are nowadays frequently confronted with massive data sets from various frontiers of scientific research. Fields such as genomics, neuroscience, finance and earth sciences have different concerns on their subject matters, but nevertheless share a common theme: they rely heavily on extracting useful information from massive data and the number of covariates p can be huge in comparison with the sample size n . In such a situation, the parameters are identifiable only when the number of the predictors that are relevant to the response is small, namely, the vector of regression coefficients is sparse. This sparsity assumption has a useful interpretation that only a limited number of variables have a prediction power on the response. To explore the sparsity, variable selection techniques are needed.

Over the last ten years, there has been many exciting developments in statistics and machine learning on variable selection techniques for ultra-high dimensional feature space. They can basically be classified into two classes: penalized likelihood and screening. Penalized likelihood techniques are well known in statistics: Lasso [29], smoothly clipped absolute deviation (SCAD) [9] or other folded concave regularization methods ([11], [30]), and Dantzig selector ([3], [4]), among others. These techniques select variables and estimate parameters simultaneously by solving a high-dimensional optimization problem. Despite the fact that various efficient algorithms have been proposed ([8], [20], [21]), statisticians and machine learners still face huge computational challenges when the number of variables is in tens of thousands of

dimensions or higher. This is particularly the case as we are entering into the era of “Big Data” in which both sample size and dimensionality are large.

The idea of feature screening came along as high-dimensional data were collected in modern technology. It was aimed at dealing with the challenges of computational expediency, statistical accuracy, and algorithmic stability because of high dimensionality. [10] proposed the sure independence screening (SIS) and showed that the Pearson correlation ranking procedure possessed a sure screening property for linear regression with Gaussian predictors and responses. However, two potential issues might arise with the screening procedures. First, an important predictor that is marginally uncorrelated but jointly correlated with the response cannot be picked by SIS and thus will not enter the estimation model. Second, this procedure works only for linear models, and performance is very unstable in other nonlinear models.

To address the first issue, an iterative SIS (ISIS) was proposed in [10] as an extension of SIS. The ISIS works as follows. In the first step, we select a subset of variables using an SIS-based model selection model. Then we have an n -vector of residuals from regressing the response over the variables selected in the first step. In the next step, we treat those residuals as the new response and apply the same method as in the previous step to the remaining variables. From the discussion above, the ISIS use a residual based approach to circumvent the problem but the idea of conditional screening has never been formally developed. To address the second issue, a new feature screening procedure for high-dimensional data based on distance correlation, named DC-SIS, was presented in [18]. DC-SIS retained the sure screening property of the SIS and additionally possessed new advantages of handling grouped predictors and multivariate responses by using distance correlation. Moreover, because distance correlation was applicable to arbitrary distributions, DC-SIS could also be used for screening features without specifying a regression model between the response and the predictors and thus was robust to model mis-specification.

In this part of the thesis, we propose an iterative feature screening procedure based on partial distance correlation which is introduced in Chapter 3. To the best of our knowledge, this is the first time that the partial distance correlation is applied to feature screening. We demonstrate our method through simulations and real examples.

The rest of the chapter is organized as follows: In Section 4.2, we give introduction to some preliminaries of partial distance covariance. In Section 4.3, we describe the iterative feature screening procedure based on partial distance correlation. In Section 4.4, we demonstrate the performance of the proposed feature screening procedure by simulations and real-data example.

4.2 *Some preliminaries*

4.2.1 Partial distance covariance

Before introducing the innovative feature screening procedure, we first review the related important properties of partial distance correlation. Let $(x, y, z) = \{(x_i, y_i, z_i), i = 1, \dots, n\}$ be a random sample observed from the joint distribution of (X, Y, Z) . The sample partial distance covariance is defined by

$$\begin{aligned} pdCov(X, Y; Z) &= (P_{z^\perp}(x) \cdot P_{z^\perp}(y)) \\ &= \frac{1}{n(n-3)} \sum_{i \neq j} (P_{z^\perp}(x))_{i,j} (P_{z^\perp}(y))_{i,j}. \end{aligned} \quad (135)$$

where

$$P_{z^\perp}(x) = \tilde{A} - \frac{(\tilde{A} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})} \tilde{C}, \quad P_{z^\perp}(y) = \tilde{B} - \frac{(\tilde{B} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})} \tilde{C}$$

denote the orthogonal projection of $\tilde{A}(x)$ onto $(\tilde{C}(z))^\perp$ and the orthogonal projection of $\tilde{B}(y)$ onto $(\tilde{C}(z))^\perp$, respectively. Here \tilde{A}, \tilde{B} , and \tilde{C} are elements of \mathcal{H}_n corresponding to samples x, y and z , where $\mathcal{H}_n = \{\tilde{A} : A \in \mathcal{S}_n\}$ contains the \mathcal{U} -centered distance matrix.

Since $pdCov$ is defined as the inner product (135) of two \mathcal{U} -centered matrices, and (unbiased squared) distance covariance is computed as inner product (122), then a

natural question is raised: are matrices $P_{z^\perp}(x)$ and $P_{z^\perp}(y)$ the \mathcal{U} -centered Euclidean distance matrices of samples of random vectors U and V , respectively? If so, then the sample partial distance covariance (135) is distance covariance of U and V , as defined by (122).

For every sample $x = \{x_1, \dots, x_n\}, x_i \in \mathbb{R}^p$ of random vector X , there is a \mathcal{U} -centered matrix $\tilde{A} = \tilde{A}(x)$ in \mathcal{H}_n . Conversely, given an arbitrary element H of \mathcal{H}_n , does there exist a configuration of points $u = \{u_1, \dots, u_n\}$ in some Euclidean space \mathbb{R}^q , for some $q \geq 1$, such that the \mathcal{U} -centered Euclidean distance matrix of sample u is exactly equal to the matrix H ?

The following theorem, which is the Theorem 2 in [27], proves that the answer is affirmative.

Theorem 4.2.1. *Let H be an arbitrary element of the Hilbert space \mathcal{H}_n of \mathcal{U} -centered distance matrix. Then there exists a sample v_1, \dots, v_n in a Euclidean space of dimension at most $n - 2$, such that \mathcal{U} -centered distance matrix of v_1, \dots, v_n is exactly equal to H .*

[26] and [28] showed that the distance correlation of two random vectors equals to zero if and only if these two random vectors are independent. Furthermore, the distance correlation of two univariate normal random variables is a strictly increasing function of the absolute value of the Pearson correlation of these two normal random variables. These two remarkable properties together with the definition of partial distance correlation motivate us to use the partial distance correlation for iterative feature screening in high-dimensional data.

For the regression model where the response and predictors all follow normal distribution, some important variables that are weakly correlated with the response through their associations with Z will not survive in SIS or DC-SIS because of the low marginal Pearson correlation or distance correlation. However, after variables in

Z enter the model as conditional variable, we have

$$pdCov(X, Y; Z) = (P_{z^\perp}(x) \cdot P_{z^\perp}(y)) = (P_{z^\perp}(x) \cdot \tilde{B}(y)) = DC(U, Y).$$

Here $pdCov(X, Y; Z)$ equals to the distance covariance of U and Y , where U is a random vector such that the U -centered distance matrix of its sample is exactly equal to $P_{z^\perp}(x)$, and U is independent of Z . Therefore, if X is marginally weakly correlated with Y purely due to the presence of Z in conditional set, it now should be correlated with the response. This addresses the first issue, that an important predictor that is marginally uncorrelated but jointly correlated with the response cannot be picked by SIS or DC-SIS. Meanwhile, partial distance correlation has the advantage of capturing nonlinear dependence. This virtue makes the proposed procedure robust to model misspecification and therefore helps to solve the second issue, that SIS may fail in nonlinear models.

4.3 An Iterative Screening Procedure

We prove the following theorem which will be implemented as a check for stopping for our feature screening procedure.

Theorem 4.3.1. *Let $(x, y, z) = \{(x_i, y_i, z_i), i = 1, \dots, n\}$ be a random sample observed from the joint distribution of (X, Y, Z) , and Let U be the random vector such that the “ \mathcal{U} -centered” Euclidean distance matrix of its sample is $P_{z^\perp}(x)$. If U is independent of Y , then*

$$pdCov(X, Y; Z) = 0 \tag{136}$$

where $pdCov$ is the sample partial distance covariance defined in (135)

Proof. Let $\tilde{A}(x)$, $\tilde{B}(y)$, and $\tilde{C}(z)$ denote the \mathcal{U} -centered distance matrix of x , y , and z . U is independent of Y is equivalent to the fact that distance covariance of U and Y is zero,

$$(P_{z^\perp}(x) \cdot \tilde{B}(y)) = 0.$$

The above inner product is defined as in (122). Since $P_{z^\perp}(x)$ denotes the orthogonal projection of $\tilde{A}(x)$ onto $\tilde{C}(z)^\perp$, then we have

$$(P_{z^\perp}(x) \cdot \tilde{C}(z)) = 0.$$

Therefore, we conclude

$$\begin{aligned} pdCov(X, Y; Z) &= (P_{z^\perp}(x) \cdot P_{z^\perp}(y)) \\ &= (P_{z^\perp}(x) \cdot \tilde{B}(y)) - \frac{\tilde{B}(y) \cdot \tilde{C}(z)}{\tilde{C}(z) \cdot \tilde{C}(z)} (P_{z^\perp}(x) \cdot \tilde{C}(z)) \\ &= 0. \end{aligned}$$

□

In reality, the test for zero partial distance covariance is implemented as a permutation test, and the test statistic is defined as in (126). The sample indices of the X sample are randomized for each replication to obtain the sampling distribution of the test statistics under the null hypothesis.

Specifically, in each permutation test, $R = 999$ replicates are generated and the estimated p-value is computed as

$$\hat{p} = \frac{1 + \sum_{k=1}^R I(T^{(k)} \geq T_0)}{1 + R},$$

where $I(\cdot)$ is the indicator function, T_0 is the observed value of the test statistic, and $T^{(k)}$ is the statistic for the k th sample. The test is rejected at significance level α if $\hat{p} \leq \alpha$.

In this section, we propose an iterative screening procedure built upon partial distance correlation. Let \mathbf{y} be the response vector, and $\mathbf{x} = (X_1, \dots, X_p)^T$ be the predictor vector. In the initial step, the first variable to enter the model is the variable X_j for which distance covariance $V_{X_j, \mathbf{y}}$ with response \mathbf{y} is largest. After the initial step, we have a model with one predictor X_j , and we compute $pdCov(X_k, \mathbf{y}; X_j)$ with the

fast algorithm proposed in Chapter 3, for the variables $\{X_k\}_{k \neq j}$, then select the variable X_k for which $pdCov(X_k, \mathbf{y}; X_j)$ is largest. Then continue, at each step computing $pdCov(X_k, \mathbf{y}; Z)$ for every X_k not yet in the vector of predictors Z currently in the model. Then the variable entered next is the one that maximizes $pdCov(X_k, \mathbf{y}; Z)$. The procedure takes the following steps, and we denote this procedure as PDCOV method.

- (1) Calculate the marginal distance covariances for $X_j, j = 1, \dots, p$ with the response. The variable X_k for which the marginal distance correlation is the largest will enter the model as predictor variable, and set $Z = \{X_k\}$.
- (2) Calculate the partial distance covariances $pdCov(X_j, \mathbf{y}; Z)$ for $X_j \notin Z$. The variable X_k for which the partial distance correlation is the largest will enter the model as predictor variable, and update $Z = Z \cup \{X_k\}$.
- (3) Repeat Step (2) until the zero partial distance covariance test for the selected variable is rejected at 5% significance level, i.e., $\hat{p} \leq 0.05$.

4.4 *Simulations Results*

In this section, we assess the performance of the PDCOV procedure by Monte Carlo simulation. Our simulation studies were conducted using Matlab code.

Example 1. The aim of this example is to examine the performance of the PDCOV in the situation where the conditions of SIS fail. More specifically, an important predictor that is marginally uncorrelated but jointly correlated with the response cannot be picked by SIS and thus will not enter the estimation model. We evaluate the methods by counting the frequencies that the selected models include all the variables in the true model, namely the ability of correctly screening “unimportant” variables.

Both linear and nonlinear models are tested in this example.

$$(1.a:) Y = 5X_1 + 5X_2 + 5X_3 - 15\rho X_4 + \epsilon,$$

$$(1.b:) Y = 5X_1 + 5X_2 + 5X_3 - \frac{15\rho}{\sqrt{2}}\mathbf{1}_{\{X_4 < 0\}} + \epsilon,$$

where X_1, \dots, X_p are p predictors and $\epsilon \in \mathcal{N}(0, 1)$ is noise that is independent of the predictors, and $\mathbf{1}_{\{X_4 < 0\}}$ is an indicator function. In the simulation, a sample of X_1, \dots, X_p with size n was drawn from a multivariate normal distribution $\mathcal{N}(0, \Sigma)$ whose covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$ has entries $\sigma_{ii} = 1, i = 1, \dots, p$ and $\sigma_{ij} = \rho, i \neq j$. In both models, the variable X_4 is uncorrelated with the response Y . Therefore, SIS ([10]) cannot pick up the true model except by chance.

We considered six such cases characterized by (p, n, ρ) with $p = 500, 1000$, $n = 50, 70, 100$ and ρ is fixed at 0.5. We simulated 200 data sets for each case. In Table 11, we report the percentages of SIS, ISIS, and PDCOV that include the true model of four variables. In this simulation example, PDCOV always picks all true variables in both linear and nonlinear models, while ISIS performs slightly worse in nonlinear model. This demonstrates that PDCOV can effectively handle the first problem that an important predictor which is marginally uncorrelated but jointly correlated with the response cannot be picked by SIS. Furthermore, in Table 12, we also report the average number of iterations in ISIS and PDCOV that are taken in order to include all the variables in the true model.

Example 2. The aim of this example is to examine the performance of the PDCOV in the situation where the model is nonlinear. In this example, the finite sample performance of the PDCOV is evaluated through the criteria \mathcal{L} : the minimum number of iterations to include all active predictors. We redo the simulations as in [18] with sample size $n = 200$. We generate $\mathbf{x} = (X_1, X_2, \dots, X_p)^T$ from normal distribution with zero mean and covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$, and the error term ϵ from the standard normal distribution $\mathcal{N}(0, 1)$. Two covariance matrices are considered to assess the performance of the PDCOV and the DC-SIS: (1) $\sigma_{ij} = 0.8^{|i-j|}$ and (2) $\sigma_{ij} = 0.5^{|i-j|}$. Note that a covariance matrix with entries $\sigma_{ij} = \rho^{|i-j|}, 0 < \rho < 1$,

Table 11: Results of simulated example I: accuracy of SIS, ISIS, and PDCOV in including the true model $\{X_1, X_2, X_3, X_4\}$

p	model	method	Results for the following value of n		
			$n = 50$	$n = 70$	$n = 100$
500	(a)	SIS	0	0	0
		ISIS	1	1	1
		PDCOV	1	1	1
	(b)	SIS	0	0	0
		ISIS	0.95	0.97	0.99
		PDCOV	1	1	1
1000	(a)	SIS	0	0	0
		ISIS	1	1	1
		PDCOV	1	1	1
	(b)	SIS	0	0	0
		ISIS	0.83	0.87	0.87
		PDCOV	1	1	1

Table 12: Results of simulated example I: average number of iterations in ISIS and PDCOV are taken to include the true model $\{X_1, X_2, X_3, X_4\}$

p	model	method	Average number of iterations for the following value of n		
			$n = 50$	$n = 70$	$n = 100$
500	(a)	ISIS	5	5	4
		PDCOV	20	18	16
	(b)	ISIS	7	7	5
		PDCOV	25	21	19
1000	(a)	ISIS	5	5	5
		PDCOV	23	20	19
	(b)	ISIS	8	8	7
		PDCOV	32	29	24

enjoys a known Cholesky decomposition: $\Sigma = R^T R$, where $R = (r_{ij}) \in \mathbb{R}^{p \times p}$, $r_{ij} = 0$, if $j < i$, and $r_{1j} = \rho^{j-1}$, $r_{ij} = c \cdot \rho^{j-i}$, for $i \geq 2$ and $j \geq i$, $c^2 + \rho^2 = 1$. In our simulations, we take advantage of this known decomposition. The dimension p varies from 2000 to 5000. Each experiment was repeated 500 times, and the performance is evaluated through the criteria \mathcal{L} : the minimum number of iterations to include all active predictors. We report the 5%, 25%, 50%, 75%, and 95% quantiles of \mathcal{L} out of 500 replications.

The \mathcal{L} is used to measure the model complexity of the resulting model of an underlying iterative screening procedure. The smaller the minimum number of iterations the \mathcal{L} is, the better the screening procedure is.

In this example, we generate the response from the following four models:

$$(2.a): Y = c_1\beta_1X_1 + c_2\beta_2X_2 + c_3\beta_3\mathbf{1}_{\{X_{12}<0\}} + c_4\beta_4X_{22} + \varepsilon,$$

$$(2.b): Y = c_1\beta_1X_1X_2 + c_3\beta_3\mathbf{1}_{\{X_{12}<0\}} + c_4\beta_4X_{22} + \varepsilon,$$

$$(2.c): Y = c_1\beta_1X_1X_2 + c_3\beta_3\mathbf{1}_{\{X_{12}<0\}}X_{22} + \varepsilon,$$

$$(2.d): Y = c_1\beta_1X_1 + c_2\beta_2X_2 + c_3\beta_3\mathbf{1}_{\{X_{12}<0\}} + \exp(c_4|X_{22}|)\varepsilon,$$

where $\mathbf{1}_{\{X_{12}<0\}}$ is an indicator function.

The regression functions $E(Y|x)$ in models (2.a)-(2.d) are all nonlinear in X_{12} . In addition, models (2.b) and (2.c) contain an interaction term X_1X_2 , and model (2.d) is heteroscedastic. Following [18], we choose $\beta_j = (-1)^U(a + |Z|)$ for $j = 1, 2, 3$, and 4, where $a = 4 \log n / \sqrt{n}$, $U \sim \text{Bernoulli}(0.4)$ and $Z \sim \mathcal{N}(0, 1)$. We set $(c_1, c_2, c_3, c_4) = (2, 0.5, 3, 2)$ in this example to be consistent with the experiments in [18]: challenging the feature screening procedures under consideration. For each step of the iterative screening procedure, we compute the associated utility between each predictor X_k and the response Y . That is, we regard $\mathbf{x} = (X_1, \dots, X_p)^T \in \mathbb{R}^p$ as the predictor vector in this example.

Table 13: The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum number of iterations \mathcal{L} out of 500 replications in Example 2

	\mathcal{L}				
	PDCOV				
Model	5%	25%	50%	75%	95%
Case 1: $p = 2000$ and $\sigma_{ij} = 0.5^{ i-j }$					
(2.a)	24	24	26	30	32
(2.b)	25	27	28	32	44
(2.c)	35	37	37	39	41
(2.d)	37	40	45	70	301
Case 2: $p = 2000$ and $\sigma_{ij} = 0.8^{ i-j }$					
(2.a)	25	29	34	41	56
(2.b)	24	32	38	40	57
(2.c)	29	32	37	41	64
(2.d)	36	40	55	80	423
Case 3: $p = 5000$ and $\sigma_{ij} = 0.5^{ i-j }$					
(2.a)	34	36	48	59	88
(2.b)	38	41	55	70	97
(2.c)	39	43	60	85	118
(2.d)	56	58	79	100	785
Case 4: $p = 5000$ and $\sigma_{ij} = 0.8^{ i-j }$					
(2.a)	35	40	56	73	116
(2.b)	38	42	56	84	137
(2.c)	52	64	79	92	178
(2.d)	66	71	87	138	995

Table 13 presents the simulation results for \mathcal{L} . The performances of the PDCOV are quite similar in models (2.a)-(2.d), indicating that the PDCOV has a pretty robust performance without model specification.

4.5 Real application on the Leukemia data

4.5.1 Data description

Leukemia is a group of cancers that usually begins in the bone marrow and results in high numbers of abnormal white blood cells. There are four main types of leukemia: acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL) and chronic myeloid leukemia (CML), as well as a number of less common types. In 2012 leukemia developed in 352,000 people globally and caused 265,000 deaths, and it occurs more commonly in the developed world [25]. It is the most common type of cancer in children, with three quarters of leukemia cases in children being ALL[5]. However, about 90% of all leukemias are diagnosed in adults, with AML and CLL being most common in adults. Thus, identification and classification of patients with different types of leukemia thus become critical for successful diagnosis and treatment.

The landmark study [12] represented the first demonstration that genomic approaches (in this case gene expression profiling) could be used to identify new cancer subtypes or assign tumors to known classes. In the paper, they demonstrate successful classification between acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) without previous knowledge of these classes. The initial leukemia data set they collected consists of 38 training samples (27 ALL and 11 AML) and 34 test samples (20 ALL and 14 AML), with 6,817 genes expression for each sample.

Here, we are going to demonstrate the performance of feature selection and classification of the two types of leukemia (AML and ALL) based on the partial distance covariance metric, with the same data set used in [12].

4.5.2 Partial distance covariance gene selection results based on all the observations

Our feature screening procedure on the gene expression data for the 72 total samples selected 13 genes (See Table 14), among which five were reported in the literature to be related to leukemia. Zyxin, a LIM domain protein, is identified as a critical regulator for the p53 Ser46 kinase HIPK2, which induces apoptosis in response to DNA damage. Zyxin expression is important to maintain HIPK2 protein stability, as HIPK2 expression is reduced in breast and thyroid carcinoma, and is functionally compromised by mutation in acute myeloid leukemia [7]. Microsomal Glutathione S-transferases (MGST) is a subfamily of Glutathione S-transferases (GST) which comprise a series of eukaryotic and prokaryotic phase II metabolic isozymes best known for their ability to catalyze the conjugation of the reduced form of glutathione (GSH) to xenobiotic substrates for the purpose of detoxification. [14] showed that microsomal GST-II is highly expressed in the human cell line K-562, a chronic myelogenous leukemia cell line. CD33 antigen (differentiation antigen) is a myeloid differentiation antigen with endocytic properties. It is broadly expressed on acute myeloid leukemia (AML) blasts and, possibly, some leukemic stem cells and has therefore been exploited as target for therapeutic antibodies for many years [17]. Transcription factor 3 (E2A immunoglobulin enhancer-binding factors E12/E47), also known as TCF3, is a protein that in humans is encoded by the TCF3 gene. This gene is involved in several chromosomal translocations that are associated with lymphoid malignancies including several types of leukemia [22].

Seven out of the 13 genes (in bold type in Table 14) were among the 50 genes selected in [12] based on the Pearson correlation with AML-ALL distinction. This shows that our PDCOV based iterative procedure is very effective in selecting the most important features.

Table 14: Partial distance covariance gene selection results based on all the observations. The genes in bold type are also among the 50 selected genes in [12].

Gene Ranking	Gene Description
1	Zyxin (X95735)
2	GLUTATHIONE S-TRANSFERASE, MICROSOMAL (U46499)
3	CD33 CD33 antigen (differentiation antigen) (M23197)
4	TCF3 Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47) (M31523)
5	ME491 gene extracted from H.sapiens gene for Me491/CD63 antigen (X62654)
6	GB DEF = Homeodomain protein HoxA9 mRNA (U82759)
7	Nucleoside-diphosphate kinase (Y07604)
8	DF D component of complement (adipsin) (M84526)
9	Macmarcks (HG1612-HT1612)
10	C1NH Complement component 1 inhibitor (angioedema, hereditary) (M13690)
11	MPO Myeloperoxidase (M19507)
12	CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage) (M27891)
13	Epb72 gene exon 1 (X85116)

4.5.3 Fivefold cross validation

In this section, we fit the nonlinear SVM models to the leukemia data sets. In order to reduce the number of genes, we apply the PDCOV procedure and use the 13 genes selected in Section 4.5.2. Here we applied a fivefold cross validation analysis on the training data to determine the tuning parameter. The performance of the final model is evaluated on test samples.

The implementation took the following steps.

- (1) Randomly partition the 38 training samples into five non-overlapping folds

$$F_1, \dots, F_5.$$

- (2) For $k = 1, \dots, 5$

- Take $(x_i, y_i), i \notin F_k$ as the training set and $(x_i, y_i), i \in F_k$ as the validation set.
- For each value of parameter $\theta \in \Theta$, build SVM model with the 13 genes and the two leukemia status variables based on the training set. Let \hat{f}_θ^{-k} be the estimate on the training set, and record the loss on the validation set

$$e_k(\theta) = \sum_{i \in F_k} \text{loss}(\hat{f}_\theta^{-k}(x_i), y_i).$$

- (3) For each tuning parameter value θ , compute the cross validation error

$$CV(\theta) = \frac{1}{K} \sum_{i=1}^K e_k(\theta) = \frac{1}{K} \sum_{i=1}^K \sum_{i \in F_k} \text{loss}(\hat{f}_\theta^{-k}(x_i), y_i).$$

We choose the value of tuning parameter that minimizes this cross validation error

$$\hat{\theta} = \arg \min_{\theta \in \Theta} CV(\theta).$$

- (4) We train on the whole 38 training sample using the selected parameter $\hat{\theta}$. Let $\tilde{f}_{\hat{\theta}}$ be the estimate on the training set. The prediction error is given by

$$PE = \sum_{(x,y) \in \text{Test Sample}} \text{loss}(\tilde{f}_{\hat{\theta}}(x), y).$$

The accuracy of the predictors was first tested by cross-validation on the training data set. The kernel used is the Gaussian radial basis function, which is defined as

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right).$$

Therefore, the parameter θ is a two dimensional vector (σ, C) , where C is the soft margin parameter. To avoid the local minimum of the cross-validation error, we randomly choose a set of 20 initial values to search for the optimal tuning parameter $\hat{\theta}$. As a result, 14 sets of $\hat{\theta}$ were found with zero cross-validation error.

We then build a final predictor based on the 38 training data set and assesses its accuracy on the 34 testing set. We list the 14 optimal θ s with corresponding prediction errors in Table 15 shown below. In total, the predictor made strong predictions for

Table 15: The optimal tuning parameters with corresponding prediction errors

	Optimal Parameter $\exp(\hat{\theta}) = (\exp(\hat{\sigma}) \exp(\hat{C}))$	Prediction Error
1	(1.712 6.859)	0
2	(1.375 0.253)	0.0588
3	(35.816 15.949)	0
4	(2.181 0.938)	0
5	(2.043 0.814)	0
6	(4.092 4.125)	0
7	(1.957 0.281)	0
8	(2.048 5.317)	0
9	(1.591 2.963)	0
10	(2.073 0.736)	0
11	(1.322 0.437)	0.0588
12	(2.458 0.308)	0
13	(1.388 0.467)	0.0588
14	(3.936 0.180)	0

the 34 samples, and the accuracy were 100% for 13 optimal tuning parameters and around 95% for three optimal tuning parameters. This shows the prediction strengths were quite high.

In summary, the above numerical example illustrates that the iterative feature screening procedure, based on partial distance covariance, successfully extracts the features which are highly correlated with the class distinction in high dimensional data. The resulting simplified model with selected variables also showed high prediction strength.

APPENDIX A

A STOCHASTIC CONTROL MODEL FOR FOREIGN EXCHANGE RESERVE MANAGEMENT

In this section, we construct a control policy that the drift is altered based on the threshold type stopping rules. The optimal threshold and the minimum cost were obtained by numerically optimizing the value function, which is suboptimal for our problem because the feasible solutions are chosen from a constrained subset of our policy space.

In section A.0.4, we describe the dynamics of the reserve accumulation in three drift case under the threshold stopping rule. The closed-form solution of the cost function are derived subsequently in section A.0.5.

A.0.4 Dynamic of the reserve accumulation in three drift threshold stopping rule

At time zero the reserve level is $R(0) = x_{21}$. We describe the controlled reserve process as a three stage process:

1. The reserve follows a reflected BM $R = \{R(t) : t \geq 0\}$ with parameters (μ_1, σ^2) and regulation part $L(t) = -\min[0, \min_{s \leq t} R(s)]$ until it hits level y_{12} , then it goes to stage two. $L(t)$ is the minimal amount of regulation (foreign reserves injection) necessary to keep the reserve level $R(t)$ from falling below the boundary zero up to t .
2. The reserve follows a BM $R = \{R(t) : t \geq 0\}$ with parameters (μ_2, σ^2) until it hits level y_{23} or x_{21} . If it hits level x_{21} before hitting level y_{23} , then it goes back to stage one; if it hits level y_{23} before hitting level x_{21} , then it goes to stage

three.

3. The reserve follows a BM $R = \{R(t) : t \geq 0\}$ with parameters (μ_3, σ^2) until it hits level x_{32} , then it goes to stage two.

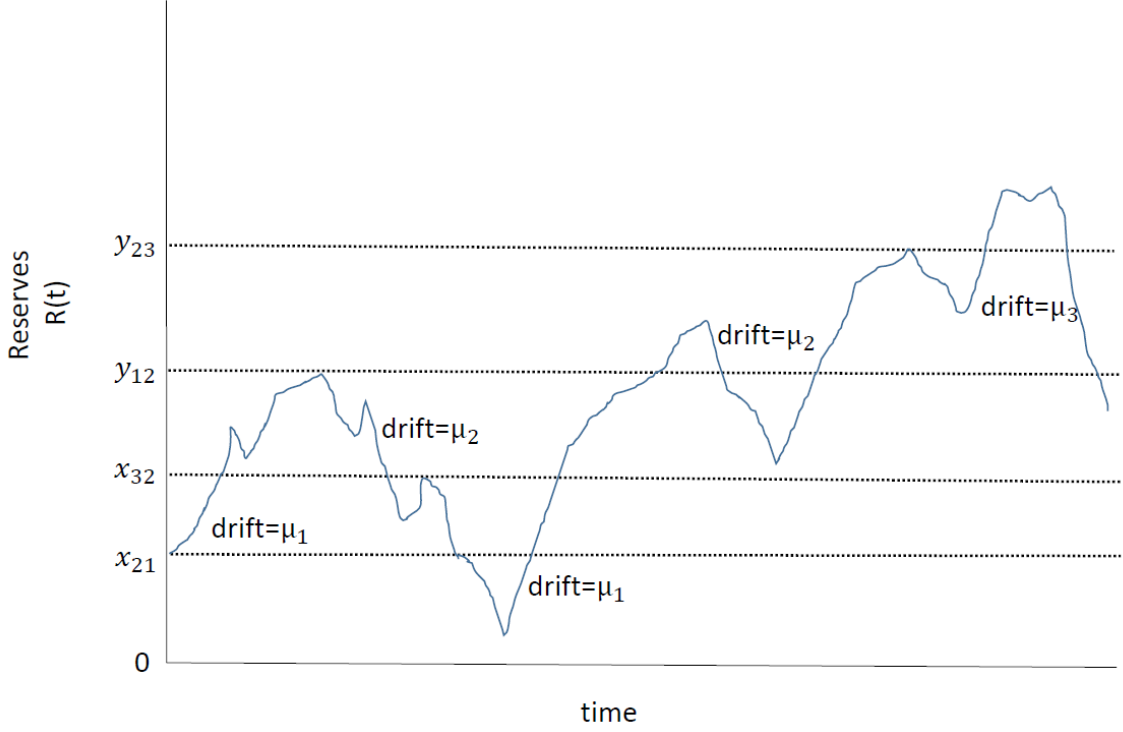


Figure 8: Dynamic of the reserve accumulation in three drift threshold stopping rule

To make it more intuitive, Figure 8 illustrates the dynamics of $R(t)$.

Having described the dynamics of drift control, we now model the costs associated with managing reserves. We identify three types of costs-the cost of holding reserves, the cost of regulation, and the cost of controlling the drift.

Let the cost of holding reserves be $hR(t)$, where h is the cost of holding \$1 of reserves per unit of time. The expected discounted cost of holding reserve is

$$A_1(\beta) = h\mathbb{E}_{x_{21},1} \left[\int_0^\infty e^{-\beta t} R(t) dt \right], \quad (137)$$

where β denotes the discounted rate and $\mathbb{E}_{z,i}[*] = \mathbb{E}[*|R(0) = z, \mu(0) = \mu_i]$.

Next we assume that there is cost k per \$1 of regulation at the boundary 0. There are an infinite and uncountable number of times that reserves hit the boundary level 0. To evaluate the regulation cost, we make use of $L(t)$ defined above. The expected discounted cost of regulation is

$$A_2(\beta) = k\mathbb{E}_{x_{21},1} \left[\int_0^\infty e^{-\beta t} dL(t) \right]. \quad (138)$$

Finally, we assume that a cost $K(\mu_i, \mu_j)$ is incurred every time the drift is switched from μ_i to μ_j and $K(\mu_i, \mu_j)$ satisfies the triangle inequality condition:

$$K(\mu_i, \mu_j) + K(\mu_j, \mu_k) > K(\mu_i, \mu_k), \quad \text{for } i \neq j, j \neq k, k \neq i. \quad (139)$$

The expected discounted cost of controlling the drift is denoted by $A_3(\beta)$.

By applying strong Markov property, we can simplify the representation of $A_i(\beta), i = 1, 2, 3$, hence get an explicit form of the total cost $M(\beta)$.

Theorem A.0.1. Define the following notation as

$$\begin{aligned}
\theta_{x_{21}}(\beta) &= \mathbb{E}_{x_{21},1}[e^{-\beta T_{x_{21},y_{12}}}], \quad T_{x_{21},y_{12}} = \inf\{t > 0 : R(t) = y_{12} | R(0) = x_{21}\}, \\
\gamma_{x_{21}}(\beta) &= \mathbb{E}_{x_{21},1} \left[\int_0^{T_{x_{21},y_{12}}} e^{-\beta t} R(t) dt \right], \\
\theta_{y_{12}}^*(\beta) &= \mathbb{E}_{y_{12},2}[e^{-\beta T_{y_{12},x_{21} \wedge y_{23}}} \mathbf{1}_{\{R(T_{y_{12},x_{21} \wedge y_{23}}) = y_{23}\}}], \\
\theta_{y_{12}^*}(\beta) &= \mathbb{E}_{y_{12},2}[e^{-\beta T_{y_{12},x_{21} \wedge y_{23}}} \mathbf{1}_{\{R(T_{y_{12},x_{21} \wedge y_{23}}) = x_{21}\}}], \\
T_{y_{12},x_{21} \wedge y_{23}} &= \inf\{t > 0 : R(t) = x_{21} \quad \text{or} \quad R(t) = y_{23} | R(0) = y_{12}\}, \\
\gamma_{y_{12}}(\beta) &= \mathbb{E}_{y_{12},2} \left[\int_0^{T_{y_{12},x_{21} \wedge y_{23}}} R(t) e^{-\beta t} dt \right], \\
\theta_{y_{23}}(\beta) &= \mathbb{E}_{y_{23},3}[e^{-\beta T_{y_{23},x_{32}}}], \\
\gamma_{y_{23}}(\beta) &= \mathbb{E}_{y_{23},3} \left[\int_0^{T_{y_{23},x_{32}}} e^{-\beta t} R(t) dt \right], \quad T_{y_{23},x_{32}} = \inf\{t > 0 : R(t) = x_{32} | R(0) = y_{23}\}, \\
\theta_{x_{32}}^*(\beta) &= \mathbb{E}_{x_{32},2}[e^{-\beta T_{x_{32},x_{21} \wedge y_{23}}} \mathbf{1}_{\{R(T_{x_{32},x_{21} \wedge y_{23}}) = y_{23}\}}], \\
\theta_{x_{32}^*}(\beta) &= \mathbb{E}_{x_{32},2}[e^{-\beta T_{x_{32},x_{21} \wedge y_{23}}} \mathbf{1}_{\{R(T_{x_{32},x_{21} \wedge y_{23}}) = x_{21}\}}], \\
T_{x_{32},x_{21} \wedge y_{23}} &= \inf\{t > 0 : R(t) = x_{21} \quad \text{or} \quad R(t) = y_{23} | R(0) = x_{32}\}, \\
\gamma_{x_{32}}(\beta) &= \mathbb{E}_{x_{32},2} \left[\int_0^{T_{x_{32},x_{21} \wedge y_{23}}} R(t) e^{-\beta t} dt \right].
\end{aligned}$$

Then we have

$$A_1(\beta) = h \cdot \frac{\gamma_{x_{21}}(\beta) + \theta_{x_{21}}(\beta) \cdot \gamma_{y_{12}}(\beta) + \frac{\theta_{y_{12}}^*(\beta) \cdot \theta_{x_{21}}(\beta)}{1 - \theta_{y_{23}}(\beta) \cdot \theta_{x_{32}}^*(\beta)} (\gamma_{y_{23}}(\beta) + \theta_{y_{23}}(\beta) \cdot \gamma_{x_{32}}(\beta))}{1 - \theta_{x_{21}}(\beta) \cdot \theta_{y_{12}}^*(\beta) - \frac{\theta_{y_{12}}^*(\beta) \cdot \theta_{y_{23}}(\beta) \cdot \theta_{x_{32}^*}(\beta) \cdot \theta_{x_{21}}(\beta)}{1 - \theta_{y_{23}}(\beta) \cdot \theta_{x_{32}}^*(\beta)}}, \quad (140)$$

$$A_2(\beta) = k \cdot \frac{E_{x_{21},1} \left[\int_0^{T_{x_{21},y_{12}}} R(t) e^{-\beta t} dL(t) \right]}{1 - \theta_{x_{21}}(\beta) \cdot \theta_{y_{12}}^*(\beta) - \frac{\theta_{y_{12}}^*(\beta) \cdot \theta_{y_{23}}(\beta) \cdot \theta_{x_{32}^*}(\beta) \cdot \theta_{x_{21}}(\beta)}{1 - \theta_{y_{23}}(\beta) \cdot \theta_{x_{32}}^*(\beta)}}, \quad (141)$$

$$\begin{aligned}
A_3(\beta) &= \frac{\theta_{x_{21}}(\beta) K(\mu_1, \mu_2) + \theta_{x_{21}}(\beta) \theta_{y_{12}^*}(\beta) K(\mu_2, \mu_1) + \theta_{x_{21}}(\beta) \theta_{y_{12}}^*(\beta) K(\mu_2, \mu_3)}{1 - \theta_{y_{23}}(\beta) \cdot \theta_{x_{32}}^*(\beta)} \\
&+ \frac{\frac{\theta_{x_{21}}(\beta) \theta_{y_{12}}^*(\beta) \theta_{y_{23}}(\beta)}{1 - \theta_{y_{23}}(\beta) \theta_{x_{32}^*}(\beta)} (K(\mu_3, \mu_2) + \theta_{x_{32}^*}(\beta) K(\mu_2, \mu_1) + \theta_{x_{32}}^*(\beta) K(\mu_2, \mu_3))}{1 - \theta_{y_{23}}(\beta) \cdot \theta_{x_{32}}^*(\beta)}. \quad (142)
\end{aligned}$$

Proof. Because of the Markov property, we can write the three costs associated with reserve management-the holding cost, the regulation cost, and the cost of changing the drifts, as the following:

$$\begin{aligned} \mathbb{E}_{x_{21,1}} \left[\int_0^\infty e^{-\beta t} R(t) dt \right] &= \mathbb{E}_{x_{21,1}} \left[\int_0^{T_{x_{21},y_{12}}} e^{-\beta t} R(t) dt \right] \\ &\quad + \mathbb{E}_{x_{21,1}} \left[e^{-\beta T_{x_{21},y_{12}}} \right] \cdot \mathbb{E}_{y_{12,2}} \left[\int_0^\infty e^{-\beta t} R(t) dt \right], \end{aligned} \quad (143)$$

$$\begin{aligned} \mathbb{E}_{y_{12,2}} \left[\int_0^\infty e^{-\beta t} R(t) dt \right] &= \mathbb{E}_{y_{12,2}} \left[\int_0^{T_{y_{12},x_{21} \wedge y_{23}}} e^{-\beta t} R(t) dt \right] \\ &\quad + \mathbb{E}_{y_{12,2}} \left[e^{-\beta T_{y_{12},x_{21} \wedge y_{23}}} \cdot 1_{\{R(T_{y_{12},x_{21} \wedge y_{23}})=x_{21}\}} \right] \\ &\quad \cdot \mathbb{E}_{x_{21,1}} \left[\int_0^\infty e^{-\beta t} R(t) dt \right] \\ &\quad + \mathbb{E}_{y_{12,2}} \left[e^{-\beta T_{y_{12},x_{21} \wedge y_{23}}} \cdot 1_{\{R(T_{y_{12},x_{21} \wedge y_{23}})=y_{23}\}} \right] \\ &\quad \cdot \mathbb{E}_{y_{23,3}} \left[\int_0^\infty e^{-\beta t} R(t) dt \right], \end{aligned} \quad (144)$$

$$\begin{aligned} \mathbb{E}_{x_{32,2}} \left[\int_0^\infty e^{-\beta t} R(t) dt \right] &= \mathbb{E}_{x_{32,2}} \left[\int_0^{T_{x_{32},x_{21} \wedge y_{23}}} e^{-\beta t} R(t) dt \right] \\ &\quad + \mathbb{E}_{x_{32,2}} \left[e^{-\beta T_{x_{32},x_{21} \wedge y_{23}}} 1_{\{R(T_{x_{32},x_{21} \wedge y_{23}})=x_{21}\}} \right] \\ &\quad \cdot \mathbb{E}_{x_{21,1}} \left[\int_0^\infty e^{-\beta t} R(t) dt \right] \\ &\quad + \mathbb{E}_{x_{32,2}} \left[e^{-\beta T_{x_{32},x_{21} \wedge y_{23}}} 1_{\{R(T_{x_{32},x_{21} \wedge y_{23}})=y_{23}\}} \right] \\ &\quad \cdot \mathbb{E}_{y_{23,3}} \left[\int_0^\infty e^{-\beta t} R(t) dt \right]. \end{aligned} \quad (145)$$

Solving the system of linear equations, we can easily get (140), (141), and (142). \square

Adding together the three costs, the total expected discounted cost of managing reserves is therefore

$$M(\beta) = A_1(\beta) + A_2(\beta) + A_3(\beta). \quad (146)$$

This completes the description of the drift control model of international reserves.

A.0.5 Solving the threshold drift control model

In this section, we compute the total expected discounted cost, $M(\beta)$ in (146). To do so, we need to derive explicit solutions for the functions

$$\theta_{x_{21}}(\beta), \quad \theta_{y_{12}}^*(\beta), \quad \theta_{y_{12}^*}(\beta), \quad \theta_{y_{23}}(\beta), \quad \theta_{x_{32}}^*(\beta), \quad \theta_{x_{32}^*}(\beta), \quad (147)$$

$$\gamma_{x_{21}}(\beta), \quad \gamma_{y_{12}}(\beta), \quad \gamma_{y_{23}}(\beta), \quad \gamma_{x_{32}}(\beta). \quad (148)$$

Define

$$m_i^+ = \frac{1}{2} \left(-\frac{2\mu_i}{\sigma^2} + \sqrt{\left(\frac{2\mu_i}{\sigma^2}\right)^2 + 4\frac{2\beta}{\sigma^2}} \right) > 0, \quad i = 1, 2, 3,$$

$$m_i^- = \frac{1}{2} \left(-\frac{2\mu_i}{\sigma^2} - \sqrt{\left(\frac{2\mu_i}{\sigma^2}\right)^2 + 4\frac{2\beta}{\sigma^2}} \right) < 0, \quad i = 1, 2, 3.$$

The following lemma tells us how to explicitly calculate the functions in (147) and (148):

Lemma A.0.2.

$$\begin{aligned} \theta_{x_{21}}(\beta) &= \frac{-m_1^+ e^{m_1^- x_{21}} + m_1^- e^{m_1^+ x_{21}}}{-m_1^+ e^{m_1^- y_{12}} + m_1^- e^{m_1^+ y_{12}}}, \\ \theta_{y_{12}^*}(\beta) &= \frac{e^{m_2^+ y_{23} + m_2^- y_{12}} - e^{m_2^+ y_{12} + m_2^- y_{23}}}{e^{m_2^+ y_{23} + m_2^- x_{21}} - e^{m_2^+ x_{21} + m_2^- y_{23}}}, \\ \theta_{y_{12}}^*(\beta) &= \frac{e^{m_2^+ x_{21} + m_2^- y_{12}} - e^{m_2^+ y_{12} + m_2^- x_{21}}}{e^{m_2^+ x_{21} + m_2^- y_{23}} - e^{m_2^+ y_{23} + m_2^- x_{21}}}, \\ \theta_{y_{23}}(\beta) &= e^{m_3^- (y_{23} - y_{12})}, \\ \theta_{x_{32}^*}(\beta) &= \frac{e^{m_2^+ y_{23} + m_2^- x_{32}} - e^{m_2^+ x_{32} + m_2^- y_{23}}}{e^{m_2^+ y_{23} + m_2^- x_{21}} - e^{m_2^+ x_{21} + m_2^- y_{23}}}, \\ \theta_{x_{32}}^*(\beta) &= \frac{e^{m_2^+ x_{21} + m_2^- x_{32}} - e^{m_2^+ x_{32} + m_2^- x_{21}}}{e^{m_2^+ x_{21} + m_2^- y_{23}} - e^{m_2^+ y_{23} + m_2^- x_{21}}}, \end{aligned} \quad (149)$$

$$\begin{aligned}
\mathbb{E}_{x_{21},1} \left[\int_0^{T_{x_{21},y_{12}}} R(t) e^{-\beta t} dL(t) \right] &= \frac{e^{m_1^- x_{21} + m_1^+ y_{12}} - e^{m_1^+ x_{21} + m_1^- y_{12}}}{-m_1^- e^{m_1^+ y_{12}} + m_1^+ e^{m_1^- y_{12}}}, \\
\gamma_{x_{21}}(\beta) &= \frac{\left(x_{21} - y_{12} \theta_{x_{21}}(\beta) + \mathbb{E}_{x_{21},1} \left[\int_0^{T_{x_{21},y_{12}}} R(t) e^{-\beta t} dL(t) \right] \right) \beta + \mu_1 (1 - \theta_{x_{21}}(\beta))}{\beta^2}, \\
\gamma_{y_{23}}(\beta) &= \frac{(y_{23} - x_{32} \theta_{y_{23}}(\beta)) \beta + \mu_3 (1 - \theta_{y_{23}}(\beta))}{\beta^2}, \\
\gamma_{y_{12}}(\beta) &= \frac{(y_{12} - x_{21} \theta_{y_{12}^*}(\beta) - y_{23} \theta_{y_{12}}^*(\beta)) \beta + \mu_2 (1 - \theta_{y_{12}^*}(\beta) - \theta_{y_{12}}^*(\beta))}{\beta^2}, \\
\gamma_{x_{32}}(\beta) &= \frac{(x_{32} - x_{21} \theta_{x_{32}^*}(\beta) - y_{23} \theta_{x_{32}}^*(\beta)) \beta + \mu_2 (1 - \theta_{x_{32}^*}(\beta) - \theta_{x_{32}}^*(\beta))}{\beta^2}. \tag{150}
\end{aligned}$$

Proof. The proof is similar to the proof in the Appendix in [2]. □

APPENDIX B

PARTIAL DISTANCE COVARIANCE: A FAST COMPUTING ALGORITHM

B.0.6 Proof of Lemma 3.3.1

Proof.

$$\begin{aligned}
 a_{i.} &= \sum_{l=1}^n a_{i,l} = \sum_{l=1}^n |x_i - x_l| \\
 &= \sum_{x_l < x_i} (x_i - x_l) + \sum_{x_l > x_i} (x_l - x_i) \\
 &= x_i \left(\sum_{x_l < x_i} 1 - \sum_{x_l > x_i} 1 \right) - \sum_{x_l < x_i} x_l + \sum_{x_l > x_i} x_l.
 \end{aligned}$$

It is easy to verify that

$$\sum_{x_l > x_i} 1 = n - 1 - \alpha_i^x,$$

and

$$\sum_{x_l > x_i} x_l = x_{.} - x_i - \beta_i^x.$$

Taking into account the above two equations, we have

$$\begin{aligned}
 a_{i.} &= (2\alpha_i^x - n + 1)x_i - \beta_i^x + x_{.} - x_i - \beta_i^x \\
 &= x_{.}(2\alpha_i^x - n)x_i - 2\beta_i^x,
 \end{aligned}$$

which is (129). □

B.0.7 Proof of Lemma 3.3.4

Proof. We have

$$\begin{aligned}
\sum_{i \neq j} a_{ij} b_{ij} &= \sum_{i \neq j} |x_i - x_j| \cdot |y_i - y_j| \\
&= \sum_{i=1}^n \sum_{j: j \neq i} (x_i y_i + x_j y_j - x_i y_j - x_j y_i) S_{ij} \\
&= \sum_{i=1}^n \left[x_i y_i \sum_{j: j \neq i} + \sum_{j: j \neq i} x_j y_j S_{ij} - x_i \sum_{j: j \neq i} y_j S_{ij} - y_i \sum_{j: j \neq i} x_j S_{ij} \right]
\end{aligned}$$

Per the definition of $\gamma_i(\{\dots\})$, one can verify that the above equates to (134). \square

B.0.8 Proof of Lemma 3.3.5

Proof. Without loss of generality (WLOG), we assume that $x_1 < x_2 < \dots < x_n$. We have

$$\begin{aligned}
\gamma_i(\{c_j\}) &= \sum_{j: j \neq i} c_j S_{ij} \\
&= \sum_{j: j > i, y_j > y_i} c_j + \sum_{j: j < i, y_j < y_i} c_j - \sum_{j: j > i, y_j < y_i} c_j - \sum_{j: j < i, y_j > y_i} c_j.
\end{aligned}$$

Note that we can verify the following equations:

$$\begin{aligned}
\sum_{j: j < i, y_j < y_i} c_j + \sum_{j: j > i, y_j < y_i} c_j &= \sum_{j: y_j < y_i} c_j, \\
\sum_{j: j < i, y_j < y_i} c_j + \sum_{j: j < i, y_j > y_i} c_j &= \sum_{j < i} c_j, \\
\sum_{j: j > i, y_j > y_i} c_j + \sum_{j: j < i, y_j < y_i} c_j + \sum_{j: j > i, y_j < y_i} c_j + \sum_{j: j < i, y_j > y_i} c_j &= c. - c_i,
\end{aligned}$$

where $c. = \sum_{j=1}^n c_j$. We can rewrite $\gamma_i(\{c_j\})$ as follows:

$$\gamma_i(\{c_j\}) = c. - c_i - 2 \sum_{j: y_j < y_i} c_j - 2 \sum_{j: j < i} c_j + 4 \sum_{j: j < i, y_j < y_i} c_j. \quad (151)$$

We will argue that the three summations on the right hand side can be implemented by $O(n \log n)$ algorithms. First, term $\sum_{j: j < i} C_j$ is a formula for partial sums. It is

known that an $O(n)$ algorithm exists, by utilizing the relation:

$$\sum_{j:j < i+1} c_j = c_i + \sum_{j:j < i} c_j.$$

Second, after sorting y_j 's at an increasing order, sums $\sum_{j:y_j < y_i} c_j$ is transferred into a partial sums sequence. Hence it can be implemented via an $O(n)$ algorithm. If QuickSort is adopted, the sorting of y_j 's is an $O(n \log n)$ algorithm.

We will argue that sums $\sum_{j:j < i, y_j < y_i} c_j, i = 1, \dots, n$, can be computed in an $O(n \log n)$ algorithm. WLOG, we assume that $y_i, i = 1, 2, \dots, n$, is a permutation of the set $\{1, 2, \dots, n\}$. WLOG, we assume that n is dyadic; i.e., $n = 2^L$, where $L \in \mathbb{N}$ or L is a nonnegative integer. For $l = 0, 1, \dots, L-1, k = 1, 2, \dots, 2^{L-l}$, we define an close interval

$$I(l, k) := [(k-1) \cdot 2^l + 1, \dots, k \cdot 2^l].$$

We then define the following function

$$s(i, l, k) := \sum_{j:j < i, y_j \in I(l, k)} c_j,$$

where $i = 1, \dots, n, l = 0, 1, \dots, L-1$, and $k = 1, 2, \dots, 2^{L-l}$.

We argue that computing the values of $s(i, l, k)$ for all i, l, k can be done in $O(n \log n)$. First of all, it is evident that

$$s(1, l, k) \equiv 0,$$

for all l, k . Suppose for all $i' \leq i$, $s(i', l, k)$'s have been computed for all l and k . For each $0 \leq l \leq L-1 < \log_2 n$, there is only one k^* , such that $y_i \in I[l, k^*]$. By the definition of $s(\cdot, \cdot, \cdot)$, we have

$$s(i+1, l, k) = \begin{cases} s(i, l, k) + c_i, & \text{if } k = k^*, \\ s(i, l, k), & \text{otherwise.} \end{cases}$$

The above dynamic programming style updating scheme needs to run for n times (i.e., for all $1 \leq i \leq n$). However, each stage requires no more than $\log_2 n$ updates. Overall, the computing for all $s(i, l, k)$ takes no more than $O(n \log n)$.

For a fixed i , $1 \leq i \leq n$, we now consider how to compute for $\sum_{j:j < i, y_j < y_i} c_j$. If $y_i = 1$, obviously we have $\sum_{j:j < i, y_j < y_i} c_j = 0$. For $y_i > 1$, there must be a unique sequence of positive integers $l_1 > l_2 > \dots > l_\tau > 0$, such that

$$y_i - 1 = 2^{l_1} + 2^{l_2} + \dots + 2^{l_\tau}.$$

Since $y_i \leq n$, we must have $\tau \leq \log_2 n$. We then define $k_\alpha, \alpha = 1, \dots, \tau$, as follows

$$\begin{aligned} k_1 &= 1, \\ k_2 &= 2^{l_1 - l_2} + 1, \\ &\vdots \\ k_\alpha &= (2^{l_1} + \dots + 2^{l_{\alpha-1}})/2^{l_\alpha} + 1, \\ &\vdots \\ k_\tau &= (2^{l_1} + \dots + 2^{l_{\tau-1}})/2^{l_\tau} + 1. \end{aligned}$$

One can then verify the following: for $2 \leq i \leq n$,

$$\sum_{j:j < i, y_j < y_i} c_j = \sum_{\alpha=1}^{\tau} s(i, l_\alpha, k_\alpha).$$

Since $\tau \leq \log_2 n$, the above takes no more than $O(\log n)$ numerical operations. Consequently, computing $\sum_{j:j < i, y_j < y_i} c_j$ for all $i, 1 \leq i \leq n$, can be done in $O(n \log n)$. \square

REFERENCES

- [1] ADELSON-VELSKII, G. and LANDIS, E. M., “An algorithm for the organization of information,” *Proceedings of the USSR Academy of Sciences*, vol. 146, pp. 263–266, 1962. (Russian) English translation by Myron J. Ricci in *Soviet Math. Doklady*, 3:1259-1263, 1962.
- [2] BAR-ILAN, A., MARION, N. P., and PERRY, D., “Drift control of international reserves,” *Journal of Economic Dynamics and Control*, vol. 31, no. 9, pp. 3110–3137, 2007.
- [3] BICKEL, P. J., RITOV, Y., and TSYBAKOV, A. B., “Simultaneous analysis of lasso and dantzig selector,” *The Annals of Statistics*, pp. 1705–1732, 2008.
- [4] CANDÈS, E. and TAO, T., “The dantzig selector: statistical estimation when p is much larger than n ,” *The Annals of Statistics*, pp. 2313–2351, 2007.
- [5] CHERNOFF, H. and PETKAU, A. J., “Optimal control of a brownian motion,” *SIAM Journal on Applied Mathematics*, vol. 34, no. 4, pp. 717–731, 1976.
- [6] CHRISTENSEN, D., “Fast algorithms for the calculation of kendalls τ ,” *Computational Statistics*, vol. 20, no. 1, pp. 51–62, 2005.
- [7] CRONE, J., GLAS, C., SCHULTHEISS, K., MOEHLENBRINK, J., KRIEGHOFF-HENNING, E., and HOFMANN, T. G., “Zyxin is a critical regulator of the apoptotic hipk2-p53 signaling axis,” *Cancer research*, vol. 71, no. 6, pp. 2350–2359, 2011.
- [8] EFRON, B., HASTIE, T., JOHNSTONE, I., and TIBSHIRANI, R., “Least angle regression,” *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [9] FAN, J. and LI, R., “Variable selection via nonconcave penalized likelihood and its oracle properties,” *Journal of the American statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [10] FAN, J. and LV, J., “Sure independence screening for ultrahigh dimensional feature space,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 5, pp. 849–911, 2008.
- [11] FAN, J. and LV, J., “Nonconcave penalized likelihood with np -dimensionality,” *Information Theory, IEEE Transactions on*, vol. 57, no. 8, pp. 5467–5484, 2011.
- [12] GOLUB, T. R., SLONIM, D. K., TAMAYO, P., HUARD, C., GAASENBEEK, M., MESIROV, J. P., COLLIER, H., LOH, M. L., DOWNING, J. R., and CALIGIURI, M. A., “Molecular classification of cancer: class discovery and class prediction by gene expression monitoring,” *science*, vol. 286, no. 5439, pp. 531–537, 1999.

- [13] HUO, X. and SZEKELY, G. J., “Fast computing for distance covariance,” *Technometrics*, 2015. Accepted.
- [14] JAKOBSSON, P. J., MANCINI, J. A., and FORD-HUTCHINSON, A. W., “Identification and characterization of a novel human microsomal glutathione s-transferase with leukotriene c4 synthase activity and significant sequence identity to 5-lipoxygenase-activating protein and leukotriene c4 synthase,” *Journal of Biological Chemistry*, vol. 271, no. 36, pp. 22203–22210, 1996.
- [15] KNIGHT, W., “A computer method for calculating kendall’s tau with ungrouped data,” *Journal of the American Statistical Association*, vol. 61, no. 314, pp. 436–439, 1966.
- [16] KONG, J., KLEIN, B. E., KLEIN, R., LEE, K. E., and WAHBA, G., “Using distance correlation and ss-anova to assess associations of familial relationships, lifestyle factors, diseases, and mortality,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 50, pp. 20352–20357, 2012.
- [17] LASZLO, G. S., ESTEY, E. H., and WALTER, R. B., “The past and future of cd33 as therapeutic target in acute myeloid leukemia,” *Blood reviews*, vol. 28, no. 4, pp. 143–153, 2014.
- [18] LI, R., ZHONG, W., and ZHU, L., “Feature screening via distance correlation learning,” *Journal of the American Statistical Association*, vol. 107, no. 499, pp. 1129–1139, 2012.
- [19] MATOGLU, M. O. and VATE, J. V., “Drift control with changeover costs,” *Operations research*, vol. 59, no. 2, pp. 427–439, 2011.
- [20] OSBORNE, M. R., PRESNELL, B., and TURLACH, B. A., “A new approach to variable selection in least squares problems,” *IMA Journal of Numerical Analysis-Institute of Mathematics and its Applications*, vol. 20, no. 3, pp. 389–404, 2000.
- [21] OSBORNE, M. R., PRESNELL, B., and TURLACH, B. A., “On the lasso and its dual,” *Journal of Computational and Graphical statistics*, vol. 9, no. 2, pp. 319–337, 2000.
- [22] PANAGOPOULOS, I., MICCI, F., THORSEN, J., HAUGOM, L., TIERENS, A., ULVMOEN, A., and HEIM, S., “A novel tcf3-hlf fusion transcript in acute lymphoblastic leukemia with at (17; 19)(q22; p13),” *Cancer genetics*, vol. 205, no. 12, pp. 669–672, 2012.
- [23] PINEAU, G., DORRUCCI, E., COMELLI, F., and LAGERBLOM, A., “The accumulation of foreign reserves,” *ECB Occasional Paper*, no. 43, 2006.
- [24] RODRIK, D., “The social cost of foreign exchange reserves,” *International Economic Journal*, vol. 20, no. 3, pp. 253–266, 2006.

- [25] STEWART, B. W. and WILD, C. P., *World Cancer Report 2014*. International Agency for Research on Cancer, 2014.
- [26] SZÉKELY, G. J. and RIZZO, M. L., “Brownian distance covariance,” *The annals of applied statistics*, vol. 3, no. 4, pp. 1236–1265, 2009.
- [27] SZEKELY, G. J. and RIZZO, M. L., “Partial distance correlation with methods for dissimilarities,” *The Annals of Statistics*, vol. 42, no. 6, pp. 2382–2412, 2014.
- [28] SZÉKELY, G. J., RIZZO, M. L., and BAKIROV, N. K., “Measuring and testing dependence by correlation of distances,” *The Annals of Statistics*, vol. 35, no. 6, pp. 2769–2794, 2007.
- [29] TIBSHIRANI, R., “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [30] ZHANG, C. H. and ZHANG, T., “A general theory of concave regularization for high-dimensional sparse estimation problems,” *Statistical Science*, vol. 27, no. 4, pp. 576–593, 2012.