

STATISTICAL MODELS FOR THE DYNAMICS OF BRAND EQUITY

Chengliang Huang

Bachelor of Science, Electronic Engineering,
Fudan University, Shanghai, 1990

Master of Business Administration
Beijing Jiaotong University, Beijing, 2007

Master of Applied Science, Electrical and Computer Engineering,
Ryerson University, Toronto, 2009

A dissertation

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Program of

Electrical and Computer Engineering

Toronto, Ontario, Canada, 2017

© Chengliang Huang 2017

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A DISSERTATION

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

Abstract

The purpose of this research is to propose statistical models, develop certain procedures/approaches needed to estimate these models, and when marketing data are available, provide insights about brand equity dynamics in marketing practice, especially firm-based brand equity. In this dissertation, two categories of models are explored. In Chapter II, autoregressive models with exogenous inputs (ARX) are proposed for brand structural analysis. These models are useful when brand values are known, and the sample size is relatively small. Another category of models, state space models, are proposed when brand values are unavailable. In Chapter III and IV, an approach or a procedure is proposed or designed to guess initial parameter values for a certain iteration algorithm. Moreover, mathematical optimization methods are introduced and integrated to estimate unknown parameters of the models for brand equity dynamics. There are at least two important findings. Firstly, the implementation of brand value structure analysis can be realized through the application of an ARX model and the assessment of a firm's brand management performance is possible. Secondly, innovative approaches must be developed to guess the starting values for iterations and to estimate parameter values of different state space models. These findings are from this innovative and contributive research. Through

brand structure analysis, a novel effort in research on brand equity dynamics, brand financial performance outcome is linked with brand equity sources, while long-term brand value is distinguished from short-term performance. The analysis helps brand managers to obtain the insights into the brand performance and the ability to focus on long-term outcomes of marketing campaigns. Moreover, innovative approaches are proposed in applying state space models for brand equity dynamics analysis. Weight least square method is used in guessing the initial parameter values for a state space model with one input series and one state series. For a state space model with two input series and two state series, as well as nonlinear constraints, a procedure is designed to guess the initial parameter values. Moreover, nonlinear mathematical optimization methods are introduced and integrated to estimate the parameter values during the implementation of the expectation-maximization algorithm.

Acknowledgements

I would like to thank my former supervisor, Dr. Xiaoping Zhang, for accepting me as his doctoral student in 2009. In August of that year, another professor at Ryerson University who had accepted me as a doctoral student informed me that I must postpone my study plan because the funding from a ministry of the Ontario government had been cancelled, affected by financial crisis. As an adult student in his forties, I did not intend to postpone my doctoral study. Moreover, my paper submitted to IEEE Sensor Journal had still not been accepted. I appreciate Dr. Zhang's belief in my research capability.

Among the graduate students in the same lab, my research area is different than other students. However, I especially appreciate the friendship with doctoral students Iris Choi, Maryam Nematollahi Arani, Roger Luo and post-doctoral Chuanlei Zhang. We shared our anxieties and happiness during our doctoral studies.

In the recent year, I am grateful for the advice given by the supervising committee. I would like to thank Dr. Soosan Beheshti and Dr. Dimitri Androutsos for providing important comments on my doctoral dissertation. I am also thankful to Dr. Miljana Horvat and Dr. Amirnaser Yazdani for chairing the supervisory committee and to Dr. Amir G. Aghdam and Dr. Richard Michon as external examiners in my exam committee.

I would like to thank my wife for being so understanding of me for many years of heavy research with limited funding. After our marriage, she did not have to work and did not have to worry about the living expenses. However, she started to take most of the responsibility in providing for the whole family. In the mean time, I am willing to offer my gratitude to my lovely and gifted daughter who brings joy and hope to me.

My parents, in their eighties, deserve my deepest and most wholehearted thanks. I am forever in debt to them. Each time I made a phone call to them, they always told me that I did not have to worry about them. They knew I was not able to spare the time and money for an international trip to visit them. I am also extremely grateful to my brother and my sister-in-law for taking care of the older members of my family while often working overtime in their schools. I perceive what sacrifices my close relatives have made and how hard they have tried so I could do the best that I can. They were the ones who make me look upward.

Acronyms

AMA	American Marketing Association
AR	Auto-Regressive (model)
ARMAX	Auto-Regressive Moving Average model with eXogenous inputs
ARX	Auto-Regressive model with eXogenous inputs
A&SP	Advertising and Sales Promotion
BLV	Brand Label Value
BOV	Brand Operation Value
BP	Brand Performance
CBBE	Consumer-Based Brand Equity
EM	Expectation-Maximization
FBBE	Firm-Based Brand Equity
HMM	Hidden Markov Models
MCMC	Markov chain Monte Carlo
MLE	Maximum Likelihood Estimation
MOM	Methods of Moments
NRMSD	Normalized Root-Mean-Square Deviation
PCA	Principal Component Analysis
PCR	Principal Component Regression
R&D	Research and Development

RMSD	Root-Mean-Square Deviation
SEM	Structural equation modeling
VSTAR	Vector Smooth Transition Auto-Regressive
WLS	Weighted Least Square
XAD	Advertising expenditures
XLR	Staff expenditures
XRD	R&D expenditures
XSF	Staff expenditures

Notation List

$\hat{\cdot}$	the guessed or estimated value of a certain variable.
α	a transfer coefficient in an ARX model, in a linear dynamic model or in the process equation of a simple state space model; a component of transfer matrix in a state space model (Sect. 4.1.1).
$\hat{\alpha}$	the estimate of α .
β	an investment effectiveness coefficient in ARX model; a component of transfer matrix in a state space model (Sect. 4.1.1).
Υ	an input coefficient in a state space model (Eq. 3.1).
γ_1	a component in the output matrix in a state space model.
γ_2	a component in the output matrix in a state space model.
$\gamma_{\varepsilon}^{(h)}$	the h th order of the auto-covariance of ε_t .
γ_{ζ}	the covariance of the error item of the output y_t .
γ_{η}	the covariance or auto-covariance of η_t .
$\gamma_{\eta}^{(h)}$	the h th autocovariance of η_t .
$\gamma_{v_1}^{(h)}$	the h th order covariance of $v_{1,t}$.
$\gamma_{v_2}^{(h)}$	the h th order covariance of $v_{2,t}$.
$\gamma_{\xi}^{(h)}$	the h th order autocovariance of ξ_t .
γ_m	the effect of the mean of output, y_t .
γ_u	the effect of the autoregressive component of the output, y_t .

$\hat{\gamma}_u$	the estimate of γ_u .
$\gamma_{\hat{u}}^{(1)}$	the first order covariance of \hat{u}_t .
$\gamma_{u_1 y}^{(-1)}$	the -1 st order cross covariance between y_t and $u_{1,t}$.
$\gamma_{u_2 y}^{(-1)}$	the -1 st order cross covariance between y_t and $u_{2,t}$.
γ_y	the covariance or autocovariance of y_t .
$\gamma_y^{(h)}$	the h th order of the auto-covariance of y_t .
Γ	the input matrix of a state space model (Eq. 4.1.1).
ε_t	the linear combination of $\zeta_{1,t}$, $\zeta_{2,t}$, η_t and w_t (Eq. 4.5.21).
ζ_t	the sum of weighted error series and observation noise in a state space model (Eq. 3.9).
$\hat{\zeta}_t$	the estimate of ζ_t .
$\zeta_{1,t}$	the Gaussian distributed stimuli component of $u_{1,t}$.
$\zeta_{2,t}$	the Gaussian distributed stimuli component of $u_{2,t}$.
η_t	the weighted sum of the error term, e_t (Sect. 4.5.2).
η_t	the sum of ξ_t and $e_{2,t}$.
θ	an unknown vector of parameters to be estimated through nonlinear optimization.
θ_1	the regression parameter for $u_{1,t-1}$ (Eq. 4.5.22).
θ_2	the regression parameter for $u_{2,t-1}$ (Eq. 4.5.22).
λ_i	the i th Lagrange multiplier (Eq. 4.4.3).
μ_0	the mean of initial state in a state space model (Eq.3.1).

μ_0	the mean of initial states in a state space model.
μ_{10}	the first component of the mean of the initial state \mathbf{x}_0 .
μ_{20}	the second component of the mean of the initial state \mathbf{x}_0 .
μ_y	the mean of y_t .
ν	the covariance of the states (Eq. 4.21a-g).
$\nu_{1,t}$	the weighted sum of $\zeta_{1,t}$.
$\nu_{2,t}$	the weighted sum of $\zeta_{2,t}$.
ξ_t	the noise component of u_t .
ξ_t	the weighted sum of $e_{1,t}$ and $e_{2,t}$ (Eq. 4.5.8).
π	the ratio of a circle's circumference to its diameter.
ρ	an autoregressive coefficient.
ρ_1	the first order autocorrelation coefficients of $u_{1,t}$.
ρ_2	the second order autocorrelation coefficients of $u_{2,t}$.
σ_0^2	the variance of x_0 .
σ_e^2	the variance of error item, e_t , in an ARX model or a dynamic linear model.
$\sigma_{\hat{\zeta}}^2$	the variance of $\hat{\zeta}_t$.
$\sigma_{\hat{u}}^2$	the variance of \hat{u}_t .
σ_w^2	the variance of w_t .
Φ	the transfer matrix of a state space model (Eq. 4.1.1).
ψ	the vector of parameters of a state space model to be estimate.
$\psi^{(j)}$	the parameter values estimated in the j th iteration of EM algorithm.

$\Omega(\Psi)$	a Lagrange function (Eq. 4.6.14).
$\Omega_0(\Psi)$	the negative log-likelihood of the initial state.
$\Omega_1(\Psi)$	the negative log-likelihood of the states other than the initial state.
$\Omega_2(\Psi)$	the negative log-likelihood of the states other than the initial state.
a	an intermediate variable (Eq. 4.5.40) in initial guessing.
Adv	advertising expense.
b	an intermediate variable (Eq. 4.5.41) in initial guessing.
BLV	brand label value.
BOV	brand operation value.
BP	brand performance.
BV	brand value.
c	a constant in an ARX model; an intermediate variable (Eq. 4.5.42) in initial guessing.
$D_y(\psi)$	the doubled negative loglikelihood as a function of the parameters to be estimated based on incomplete data, $\mathbf{y}_{1:T}$.
$D_{x,y}(\psi)$	the doubled negative loglikelihood as a function of the parameters to be estimated based on incomplete data, $\mathbf{x}_{0:T}$ and $\mathbf{y}_{0:T}$.
$E(\Omega)$	the sum of $E(\Omega_0)$, $E(\Omega_1)$ and $E(\Omega_2)$ and possibly other items.
$E(\Omega_0)$	the log-likelihood estimated from the initial state of a state space model.
$E(\Omega_1)$	the log-likelihood estimated from the other states of a state space model.
$E(\Omega_2)$	the log-likelihood estimated from the output of a state space model.

e_t	the error item in an ARX model or in a linear dynamic model, or the process error in a state space model.
\mathbf{e}_t	the vector of process noises in a state space model (Eq. 4.1.1).
$e_{1,t}$	the first component of the err vector, \mathbf{e}_t .
$e_{2,t}$	the second component of the err vector, \mathbf{e}_t .
\mathbf{H}	the design matrix or observation matrix of a state space model (Eq. 4.1.2).
$h_i(\boldsymbol{\psi})$	the constraint to the parameters to be estimated.
\mathbf{I}	an identity matrix.
j	an intermediate variate used for backward recursion in Kalman smoothing (Eq. 3.22).
\mathbf{J}	an intermediate matrix used for backward recursion in Kalman smoothing (Eq. 4.2.8).
k	the Kalman gain (Eq. 3.21a-g).
k_1	equals $\beta/(1 - \alpha)$ (Eq. 4.5.7).
k_2	a constant.
$L_{x,y}$	the Lagrange function based on complete data.
M	the sum of the product of estimated states and inputs of a state space model.
m_1	an intermediate variable (Eq. 4.5.34) in initial guessing.
m_2	an intermediate variable (Eq. 4.5.35) in initial guessing.

M_1	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.32).
M_α	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.30).
M_β	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.31).
M_σ	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.39).
$M_{\sigma 0}$	an intermediate variable regarding $x_{1,t}$ and $x_{2,t}$ in estimating parameters of a state space model (Eq. 4.6.22).
MKT	marketing spending.
$NRMSD_4$	a “normalized” measure of the differences between the guessed or estimated values, of four parameters of interest.
$NRMSD_9$	a “normalized” measure of the differences between the guessed or estimated values, of nine parameters of interest.
p	the correlation matrix of smoothed state.
P	the sum of the correlation between states of a state space model.
\mathbf{P}	the correlation matrix of the state vector.
$p(\cdot \cdot)$	a generic conditional probability density or mass function.
P_1	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.29).

P_α	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.27).
P_β	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.28).
$P_{\sigma 0}$	an intermediate variable regarding $x_{1,t}$ only in estimating parameters of a state space model (Eq. 4.6.21).
P_σ	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.38).
P_w	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.40).
Q	the variance-covariance matrix of e_t in a state space model
q_1	the variance of $e_{1,t}$.
q_2	the variance of $e_{2,t}$.
q_x	the covariance between $e_{1,t}$ and $e_{2,t}$.
$q(\psi' \psi)$	a proposal for an MCMC that explores $p(\psi' \mathbf{y}_{0:T})$.
r	the variance of w_t .
R	the variance-covariance matrix of \mathbf{w}_t in a state space model.
RD	research and development expense.
r_{perf}	brand performance ratio.
$RMSD_4$	a measure of the differences between the guessed or estimated values, of four parameters of interest.

$RMSD_9$	a measure of the differences between the guessed or estimated values, of nine parameters of interest.
s	the number of constraints.
S	the weighted sum of squared residuals (Eq. 3.12).
T	the sample size.
U	the sum of the squared inputs of a state space model.
u_t	the input in a state space model (Eq. 3.1).
\mathbf{u}_t	the vector of inputs in a state space model (Eq. 4.1.1).
$u_{1,\infty}$	the long term average of $u_{1,t}$.
$u_{2,\infty}$	the long term average of $u_{2,t}$.
$u_{1,b}$	the constant component of $u_{1,t}$ (Eq. 4.5.1).
$u_{2,b}$	the constant component of $u_{2,t}$ (Eq. 4.5.1).
$u_{1,t}$	the first component of \mathbf{u}_t .
$u_{2,t}$	the second component of \mathbf{u}_t .
U_1	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.35).
U_{γ_1}	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.33).
U_{γ_2}	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.34).

U_σ	an intermediate variable in estimating parameters of a state space model (Eq. 4.6.40).
$U_{\sigma 0}$	an intermediate variable regarding $x_{2,t}$ only in estimating parameters of a state space model.
v_{10}	the variance of the first component of the initial state \mathbf{x}_0 .
v_{20}	the variance of the second component of the initial state \mathbf{x}_0 .
v_ε	the variance and auto-covariance of ε_t .
v_{ζ_1}	the variance of $\zeta_{1,t}$.
v_{ζ_2}	the variance of $\zeta_{2,t}$.
v_η	the variance of η_t .
v_ξ	the variance of ξ_t .
v_n	the difference between v_n and σ_u^2 .
v_t^T	the variance of the smoothed estimation of state.
v_{x0}	the covariance of the first and the second component of the initial state \mathbf{x}_0 .
v_y	the variance of y_t .
V	the variance-covariance matrix of states in a state space model.
V_0	the variance-covariance matrix of initial states in a state space model.
w_t	observation noise in a state space model at time instant t .
W	the sum of the product of states and outputs of a state space model.
\mathbf{w}_t	the vector of output (observation) noises in a state space model (Eq. 4.1.2).

x_0	initial state in state space model (Eq. 3.1).
\mathbf{x}_0	the vector of initial states in a state space model.
$\overline{x_{1,0}^2}$	the correlation of $x_{1,0}$.
$\overline{x_{2,0}^2}$	the correlation of $x_{2,0}$.
$\overline{x_{1,0}x_{2,0}}$	the correlation between $x_{1,0}$ and $x_{2,0}$.
$\overline{x_{1,0}}$	the mean of $x_{1,0}$.
$\overline{x_{2,0}}$	the mean of $x_{2,0}$.
x_t	the state in a state space model (Eq. 3.1).
\mathbf{x}_t	the vector of hidden states in a state space model (Eqs. 4.1.1 and 4.1.2).
\mathbf{x}_t^T	the mean of the smoothed estimation of state.
\mathbf{x}_t^T	the state estimated through Kalman smoothing.
$\mathbf{x}_{t:T}$	$T-t+1$ samples of state \mathbf{x}_t .
\mathbf{x}_{-t}	the whole state process from \mathbf{x}_0 to \mathbf{x}_T excluding \mathbf{x}_t .
Y	the sum of the squared outputs of a state space model.
y_b	the effect of the mean of the output, y_t .
\hat{y}_b	the estimate of y_b .
y_t	the output (observation) of a state space model.
\mathbf{y}_t	the vector of outputs in a state space model (Eq. 4.1.2).
$\mathbf{y}_{t:T}$	$T-t+1$ samples of state \mathbf{y}_t .
\tilde{y}_t	the innovation item in Kalman filtering (Eq. 3.21c).
z_t	the weighted sum of input time series in a state space model.

Table of Contents

Abstract.....	iii
Acknowledgements.....	v
Acronyms.....	vii
Notation List.....	ix
Table of Contents.....	xix
List of Tables.....	xxiii
List of Figures.....	xxiv
1 Introduction.....	1
1.1 Motivation.....	1
1.2 Background.....	3
1.2.1 CBBE.....	3
1.2.2 FBBE.....	4
1.2.3 Brand persistence and brand equity drivers.....	5
1.2.4 Extant models.....	8
1.3 Proposed Models	10
1.4 Objectives.....	11
1.5 Contributions.....	14
1.5.1 Innovations and contributions.....	15
1.5.2 Challenges.....	16
1.6 Outline.....	18
2 ARX Models for Brand Value Structure Analysis.....	19
2.1 Brand Structure Analysis.....	19

2.2	Theoretical Conceptualization.....	23
2.2.1	Brand value structure.....	24
2.2.2	Modeling brand value structure.....	26
2.2.3	Brand intrinsic value.....	28
2.2.4	Brand performance ratio.....	30
2.2.5	Implications from the brand value structure analysis.....	31
2.3	Empirical Illustrations.....	33
2.3.1	Data.....	33
2.3.2	Procedures.....	37
2.3.3	Endogeneity	40
2.3.4	Results.....	40
2.4	General Discussion.....	47
2.4.1	Theoretical contributions	47
2.4.2	Managerial contributions	50
2.4.3	Methodological contributions.....	51
2.4.4	Limitations and opportunities for further research	52
2.5	Chapter Summary	53

3 A State Space Model and its Maximum Likelihood

Estimation.....	55
3.1 A State Space Model and MLE.....	55
3.2 Initial Guessing.....	58
3.2.1 Initial guessing of the system parameter and the variances.....	59
3.2.2 Initial state mean and control parameter	60
3.3 Estimation Using the EM Algorithm	61
3.3.1 The EM Algorithm.....	62
3.3.2 Kalman filtering and smoothing.....	63
3.3.3 Expected log-likelihood formulation.....	64
3.3.4 Estimation of the parameters.....	65
3.4 Data Generation and Simulation Results.....	70
3.4.1 Data Generation.....	70
3.4.2 Results.....	70
3.5 Conclusions.....	72
3.6 Chapter Summary.....	73

4 The Estimation of a State Space Model with Partly Known

Parameters and Nonlinear Constraints.....	74
4.1 A State Space Model for Brand Equity Dynamics.....	75

4.1.1	A state space model for brand equity dynamics.....	76
4.1.2	Characteristics of the parsimonious model.....	78
4.2	Existing Models: MLE.....	81
4.2.1	Kalman Filter and Kalman Smoothing.....	81
4.2.2	Mathematical optimization problem.....	82
4.2.3	Algorithm for MLE: Newton-Raphson.....	83
4.2.4	Algorithm for MLE: EM.....	84
4.3	Existing Method: Markov Chain Monte Carlo.....	87
4.3.1	Update the state of a state space model.....	88
4.3.2	Updating the parameters.....	90
4.3.3	Particle algorithm.....	91
4.3.4	Summary.....	92
4.4	Proposed Approach.....	93
4.4.1	MLE vs. MCMC.....	93
4.4.2	Mathematic problem and its characteristics.....	96
4.4.3	The EM algorithm and starting value.....	97
4.5	Guessing of the Starting Values of the Parameters.....	99
4.5.1	Investigation on the input series and the hidden states.....	100
4.5.2	Linear regression and cross covariance between input series and output series.....	102
4.5.3	Obtaining the guessed value of the unknown parameters.....	105
4.5.4	Steps for the guessing	107
4.6	MLE Using the EM Algorithm.....	108
4.6.1	“Expectation” Value for the E step.....	110
4.6.2	Mathematical optimization for the M step.....	115
4.6.3	Using first order necessary conditions to estimate μ_{10} , μ_{20} , γ_1 and γ_2	115
4.6.4	Nonlinear optimization methods.....	116
4.7	Simulation, Results and Discussion.....	118
4.7.1	Generating input series and output series.....	119
4.7.2	Guessing the parameters.....	120
4.7.3	Estimating the parameters.....	121
4.7.4	RMSD and NRMSD.....	122
4.7.5	Discussion.....	123
4.8	Conclusion and Future Research.....	125
4.9	Chapter Summary.....	128

5 Conclusion and Future Work.....130

5.1	Conclusion.....	130
5.1.1	Contribution.....	133

5.2 Future Work.....	135
Bibliography.....	137

List of Tables

1.1 Extant Statistical Models for Brand Equity Dynamics.....	8
2.1 Four types of brand management scenarios	32
2.2 Descriptive statistics and correlations.....	35
2.3 Principal component regression analysis results.....	41
2.4 Brand analysis.....	42
3.1 The parameters obtained with small sample size (Sample Size: 50).....	71
3.2 The parameters obtained with small sample size (Sample Size: 1000).....	71
4.1 State space models estimated using MLE and EM in extant literature.....	95
4.2 State space models estimated using methods based on MCMC in extant literature.....	95
4.3 Statistical indexes for input series.....	119
4.4 Regression parameters for the two-input model.....	120
4.5 Comparison among preset, guessed and estimated parameters.....	121
4.6 Comparison among preset, selected and guessed and estimated parameters.....	122

List of Figures

2.1	Decomposition of Brand Performance Outcomes	24
2.2	Brand and Brand Performance	30
2.3	Brand Value, Brand Intrinsic Value, and Brand Marketing Spending	45
4.1	Directed Acyclic Graph for the Guessing Steps.....	107

Chapter I

Introduction

1.1 Motivation

The American Marketing Association (AMA) [1] defines a brand as “a name, term, design, symbol, or any other feature that identifies one seller’s good or service as distinct from those of other sellers”. Brand equity refers to the value of a brand, which is created by the intangible qualities associated with the brand. For example, Farquhar [2] characterizes brand equity as the value added by the brand name to a product.

High brand equity relates to high market recognition and added value to a product. Accordingly, the product becomes more profitable through brand loyalty, premium pricing, lower price elasticity, lower advertising/sales ratios, and trade leverage [3]. Because it is not easy to copy brand competitors [4], brand equity also makes a brand more robust to environmental and competitive threats. For instance, Aaker (p.12) [5] indicates that the Intel Inside campaign increased its brand equity and thus enhanced Intel’s competitive

position. Consequently, those computers without an Intel microprocessor had to be sold with a price discount. This typical example indicates that brands have become one of the most valuable assets that a company has [6]. Firms of any size, in any industry, and in any type of markets, must put brand building and management as a top priority [7].

In recent years, the determination of the long-term effects of a firm's investments on marketing and research and development (R&D) has received much attention from practitioners and academics. For example, Oswald Grubel, CEO of the Credit Suisse Group, stated that "Our priorities are quite clear: we want to generate a long-term added value for our shareholders by offering outstanding service to our clients and by securing a leading position in the industry" [8]. Brand equity, as the symbol of products and services and one of the components of intangible assets, is an important measure to reflect the effects of the long-term marketing and R&D investments.

Since brand equity has been perceived as a key strategic asset that needs to be monitored and nurtured for maximum long-term performance, many firms have brand equity (BE) managers who are responsible for managing brand equity to improve marketing productivity. For brand managers to detect signals of brand equity erosion or approve branding programs [9], they need both current brand equity estimates and the dynamic tracking of brand equity over a long period, e.g., a few years. Both, i.e., estimating and tracking brand equity, however, are difficult tasks.

Our motivation in a long run is to construct dynamic mathematical models which are used in engineering research, appropriate to available data and marketing practice and able to explain brand persistence, the effect of advertising, and R&D expense.

1.2 Background

A few approaches provide brand measures. Some are from the consumer perspective, i.e., consumer-based brand equity (CBBE), and others are from the firm perspective (or product perspective), i.e., firm-based brand equity (FBBE). It is worthy of mentioning that, all these measurements from different approaches are, at best, approximations. A more complete understanding of brand equity is possible if we apply multiple approaches and cross-examine the results. Brand equity estimates based on store-level data, when combined with traditional survey-based measures, as well as firm-based estimates, can build a brand manager's confidence in formulating marketing programs to build a brand's equity. In this research, CBBE or FBBE will be chosen based on the availability of data and the possible need of marketing practice.

1.2.1 CBBE

CBBE approaches, are based on value consumers derived from brand names. Some researchers have defined this added value as the positive associations, awareness, loyalty, and perceived quality of the brand [9], or as the differential effect of brand knowledge to the marketing of a firm [10], or as the price premium that consumers are willing to pay for the brand [11]. Srinivasan [12] and Kamakura and Russell [13] isolate brand equity as a component of the overall preference not explained by objectively measured attributes.

CBBE approaches seek to map consumer minds to identify brand associations consumers have. These approaches seek to measure brand awareness, i.e., consumer ability to recall and recognize a brand, and brand image, i.e., the overall brand associations. Brands with high levels of awareness and strong, favorable and unique associations are high equity

brands [10]. Commercially, this approach is realized through consumer or expert surveys [5]. This is the case at EquiTrend and Brand Asset Evaluator. In the marketing literature, a conceptual framework for measuring CBBE using consumer surveys is offered by Keller [10] and a survey-based measure of CBBE is proposed by Park and Srinivasan [11]. Survey-based CBBE measures are less affected by market activities and other firm activities, thus managers can use them to track brand equity. However, this approach “depends on the ability of consumers to accurately report their relative brand preference”, as noted by Park and Srinivasan [11]. Again, providing exact measures is a difficult task since an accurate measurement requires both reliable measurement techniques and sound data.

1.2.2 FBBE

On the other hand, most FBBE approaches measure brands as financial assets, specifically intangible assets. The measurement of FBBE is a calculation on how much a brand is worth. For example, Neumeier [6] calculates FBBE as the residual of the subtraction of tangible assets and "measurable" intangible assets from the value of the firm, which is derived by its market capitalization. Simon and Sullivan [14] calculate FBBE as incremental cash flows attributable to branded versus unbranded products. Instead of financial market data, Mahajan et al. [15] use purchase data when a brand is sold or acquired. Aaker [5] [16] proposes a product level brand measurement through comparison between the price or market share of a no-name product to a branded product. The brand equity is the difference in price or market share. Ailawadi et al. [17] have advocated a revenue premium approach.

A widely recognized FBBE approach is from the Interbrand Group, a consulting firm. Interbrand estimates FBBE based on projected profits discounted to a present value [5] [10]. The discount rate is a subjective rate determined by Interbrand and Wall Street equity specialists and reflects the performance of a brand along different dimensions such as the risk profile, market leadership, stability and global reach of the brand [18].

According to Hanssens and Dekimpe [19], appropriate brand equity metrics should have financial relevance. But as explained later there are few models on FBBE. Consequently, in the research, FBBE are mostly considered as the measure for brand equity. However, the models we propose, especially those models in Chapter III and Chapter IV, can also be used in a CBBE scenario.

1.2.3 Brand persistence and brand equity drivers

A. Persistence of Brand Equity

Brand equity has its persistence, or inertia. Aaker [20] believes “a strong brand can withstand almost anything”. One of the causes of brand equity persistence is brand loyalty [21]. The common contention is that brand loyalty gradually decline over time. Pare and Dawes [22] point out that it not rare for a brand, as a market leader, to obtain loyalty over multiple years. The factors influencing brand loyalty include customers' perceived value, brand trust, customers' satisfaction, repeat purchase behavior, and commitment.

B. Drivers of Brand Equity

The factors affecting the evolution of brand value over time are ideally the causes of the current brand equity but not the result of brand equity. They can be classified into two categories: market strategy variables and market structure variables [23]. The market

strategy variables include the expenses on advertising, R&D, and even human resources. The market structure variables can be the concentration ratio, the different industries and the ages of firms. On the other hand, the effects of these drivers vary with industries and even countries. For example, Chen [24] finds that country, industry and firm factors will effectively increase the brand values. When brand equity dynamics of a certain firm are investigated, the market structure variables are usually omittable.

Advertising and Sales Promotion

Advertising and sales promotions (A&SP) are two central elements of marketing communication programs. Advertising is reported to have positive association with brand equity in most literature. Keller [3] suggests that advertising can increase brand equity through favorable associations, perceived quality, and user experience. In empirical research, Jedidi et al. [25] find a positive relationship between advertising and brand equity, using household purchase data. Chu and Keh [18] find advertising expenditure has positive effects on brand values using firms' financial performances. Wang et al. [26] conclude advertising has sustainable and even accumulative effects on intangible asset of a firm. Buila et al. [27] conclude that advertising expense does not impact brand associations and perceived quality but improves brand awareness.

No agreement has been achieved on the effects of promotion. Keller [3] suggests that the frequent use of price promotion may create or strengthen a "discount" association with a brand, thus diminishing its brand equity. Pauwels et al. [28] claim that price promotion has only short impact to an established brand. However, Ailawadi et al. [17] suggest that sales promotions can induce trials and thus increase penetration and brand equity. D'Astous and Jacob [29] find monetary sales promotions can prevent association

about product price from direct hurting brand equity. Buila et al. [27] also find that monetary promotions have a negative influence whereas non-monetary promotions have a positive effect on brand equity.

When both advertising and sales promotion are considered, Low and Mohr [30] find that brands allocating higher budget to advertising than to sales promotion, have more favorable consumer attitudes and stronger brand equity. Moreover, Sedaghat et al. [31] find that marketing promotional mix have positive effects on brand equity dimensions including brand loyalty, perceived quality and brand awareness. Low and Mohr [30]'s finding is based on survey data collected from 165 brand managers in the USA while Sedaghat et al. [31]'s research involves structural equation modeling (SEM) [32] [33]. Both findings are not specified to certain products.

Research & Development

For a corporate entity, R&D is to improve existing or invent new products procedures or services, or even to create new knowledge to enable discovery of new products, procedures and services. R&D is important for some firms to upgrade and/or expand a firm's operation and finally to satisfy their customers. Nowadays, R&D is of special importance for firms under high level of competition when they are trying to keep up with modern trends of technology, business and/or customer needs. "The business climate for R&D-active companies has continued to improve" [34]. Although R&D is often thought of as associated with high-tech firms which are on the cutting edge of new technology, many established consumer goods companies spend large sums of money on improving old products. For example, Gillette spends quite a bit on R&D each year in ongoing attempts to design a more effective shaver.

A few empirical researchers studied the relationship between R&D and brand equity. Simon and Sullivan [14] find that both patent share and R&D share have positive impacts on brand equity based on their research on a wide range of manufacturing industries with data from Compustat and NBER databases. Chu and Keh [18] find that R&D, together with the number of patents, has positive effects on brand values.

1.2.4 Extant models

Table 1.1 Extant Statistical Models for brand equity Dynamics

Papers	CBBE/FBBE	Statistical Models
Simon and Sullivan (1993) [14]	FBBE	Linear regression
Dillon et al. (2001) [35]	CBBE	Similar with State Space Model
Chu and Keh (2006) [18]	FBBE	Vector Autoregressive Model (Interbrand brand values are used)
Sriram and Manobar (2007) [36]	CBBE	State Space Model (Utility is used)
Shanker et al. (2008) [37]	CBBE	Linear regression (Utility is used)
Aribarg and Arora (2008) [38]	CBBE	State space model
Voleti and Ghosh (2014) [39]	CBBE	Linear regression (Revenues are used)

Other than using appropriate metrics, we also need to establish a statistical model to disentangle temporary (short-term) from persistent (long-term) effects. There are at least 400 articles published on brand equity since 2013, most of which, however, are still conceptual or about CBBE. The statistical models using FBBE in the effort to describe brand equity dynamics are rare. We have to review research papers more than 10 years old. Typical models using FBBE are represented by Simon and Sullivan's [14] linear regression model and Chu and Keh [18] 's vector regression (VAR) model. Simon and Sullivan [14] select a proxy for the FBBE using Tobin's Q [40] and developed a linear regression model to examine the effects of certain marketing variables, including advertising expenditures.

To separate the market share attributable to brand factors, they regress the observed market share on all of the factors:

$$S = b_0 + b_1 * ord + b_2 * adshr + b_3 * patshr + b_4 * rndshr + \varepsilon \quad (1.1)$$

where they use the order of market entry (*ord*) and the brand's advertising expenditures relative to its competitors' (*adshr*) as proxies for information expenditures and positioning advantage. They posit that the firm's technological advantages are related to *patshr*, the firm's share of patents relative to competitors and *rnd* share, the firm's share of R&D expenditures. The noise, ε , is the error term. The parameters, b_0 , b_1 , b_2 , b_3 and b_4 , are to be estimated. By the way, it shall be mentioned that other mathematical equations are needed for their research.

Unfortunately, the model can not reflect the persistence of brand equity thus is not able to fully represent the dynamics of brand equity. On the other hand, Chu and Keh [18], in their VAR model, relate advertising expense and R&D expense to the lagged brand valuation and others. Their model is as below:

$$\begin{aligned} \ln(BV_t) = & A_0 + \alpha_1 \ln(BV_{t-1}) + \alpha_2 \ln(NetIncome_t) + \alpha_3 \ln(Ad_{t-1}) + \\ & \alpha_4 \ln(Promo_{t-1}) + \alpha_5 \ln(R\&D_{t-1}) + \alpha_6 [\ln(Ad_{t-1})]^2 + \alpha_7 [\ln(Promo_{t-1})]^2 + \\ & \alpha_8 [\ln(R\&D_{t-1})]^2 + error, \end{aligned} \quad (1.2)$$

$$\ln(Ad_t) = B_0 + \beta_1 \ln(BV_{t-1}) + \beta_2 \ln(Ad_{t-1}) + \beta_3 \ln(Promo_t) + \beta_4 \ln(R\&D_t) + error, \quad (1.3)$$

$$\begin{aligned} \ln(Promo_t) = & \Gamma_0 + \gamma_1 \ln(BV_{t-1}) + \gamma_2 \ln(Promo_{t-1}) + \gamma_3 \ln(Ad_t) + \gamma_4 \ln(R\&D_t) + \\ & error, \end{aligned} \quad (1.4)$$

$$\ln(R\&D_t) = \Delta_0 + \delta_1 \ln(BV_{t-1}) + \delta_2 \ln(R\&D_{t-1}) + \delta_3 \ln(Ad_t) + \delta_4 \ln(Promo_t) + error, \quad (1.5)$$

where the variables are *BV* for brand value, *Ad* for advertising expenditure, *Promo* for promotion expenditure, and R&D for research and development. *NetIncome* is for a firm's net income. The *error* terms in these equations are different from each other. The parameters, $A_0, \alpha_1, \alpha_2, \dots, \alpha_8, B_0, \beta_1, \beta_2, \beta_3, \beta_4, \Gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \Delta_0, \delta_1, \delta_2, \delta_3$, and δ_4 , are to be estimated.

To apply VAR, it shall be assumed that investments must be decided proportionally with the unknown brand value and other investments. This assumption is not reasonable in the decision making on a firm's investments.

1.3 Proposed Models

To obtain a better representation of the dynamics of brand equity, we propose two types of statistical models: (1) autoregressive model with exogenous inputs (ARX), (2) state space models.

When Interbrand's brand values are used, an ARX model takes the advantages of both linear regression and VAR model. The model can reflect not only the effect of investment but also the evolvement of brand value over time. Especially, with limited sample size, it is still possible to estimate the parameters of the model.

As we know, so far state space models are used in the literature where CBBE is applied but not where FBBE is used. Considering that FBBE is an intangible asset and states in a state model are hidden values, state space model is a reasonable choice. In the meantime, the investment on marketing and R&D is treated as inputs while sales, or other brand performances, are treated as outputs.

However, the estimation of the parameters of such a state space model is challenging. In the extant literature on control engineering and time series analysis, all the parameters of the state space models are to be estimated while no exogenous inputs exist. For the identification of such a dynamic system, maximum likelihood estimation (MLE), often together with the expectation-maximization (EM) algorithm, is the popular solution. However, in marketing research, the parameters of the state space models are partly known while the exogenous input series exist. To estimate such models, Markov chain Monte Carlo (MCMC) based methods are popularly used. But the MCMC based methods used in extant literature are usually sophisticated and need additional computation and an unbounded running time. To avoid additional computation while taking the advantages of the difference between the proposed state space model and those models in marketing literature, new parameter estimation approaches are intentionally to be developed based on MLE. Considering that the task is challenging, a state space model with one input will be tried first. A state space model with two exogenous inputs, partly known parameters and nonlinear constraints will be investigated.

1.4 Objectives

We treat the brand investments as exogenous inputs. The existence of exogenous inputs is critical in classifying a system. In a dynamic system, the current output values depend not only on their earlier values but also on exogenous inputs [41]. The outputs of a system, whose exogenous inputs are not observed, are often called time series. The task of system identification is to form model through data analysis using stored input series and output series from the system.

Based on what we have investigated in Sect. 1.2, when brand values are known and sample size is not large, we intend to apply ARX models and use Interbrand brand values of the global best brands to illustrate the brand value structural analysis. Autoregressive (AR) models are conventional models for time series analysis in economy and their applications are popular. However, the estimation of ARX models shall not be confused with the application of AR models. Firstly, time series analysis is performed based on output series only while the identification of system is performed based on time varying inputs and outputs. From the viewpoint of machine learning, if only the outputs of the system can be observed, the problem is unsupervised. If both inputs and outputs are observed, the problem becomes supervised. Therefore, our research using ARX model is of system identification or supervised learning while those papers based on AR models are of time series analysis or unsupervised learning. Secondly, in the application of a model, the parameters of the model are known and the purpose of the application is to obtain outputs from known inputs using the known model. In contrast, in the estimation of a model, the outputs, together with the inputs, of the model are known. The purpose of the estimation is to obtain the unknown parameter values from known inputs and outputs. Therefore, principally the estimation of an ARX model is more complicated and challenging than the application of an AR model.

On the other hand, when brand values are unknown and a much larger sample size is available, state space models are appropriate to model the dynamics of brand equity because the brand itself is a hidden measure. As to be further introduced in Chapter III and Chapter IV, we acknowledge that state space models have been widely used in object-tracking, navigation, computer vision, econometrics and many more areas. However, in the

investigation of the dynamics of brand equity, state space models are applied to CBBE only in extant literature. Therefore, the intention to apply state space models to FBBE is innovative. Moreover, the state space models proposed, especially the models proposed in Chapter III and Chapter IV to represent brand equity dynamics, are models with time-varying exogenous inputs. Such inputs are not available in the state space models in extant literature on time series analysis. In addition, the state space model proposed in Chapter IV has nonlinear constraints and partly known parameters. As we know, the estimation of such a model has not been presented in extant literature.

However, the identification of state space models through parameters estimation is not presented yet, especially for the models with exogenous time-varying inputs, nonlinear constraints and partly known parameters. The estimation of state space model has been proved to be a task of difficulty and sophistication when the dimension of the states, the output and the inputs are increasing. To make the work less challenging, we need to start from state space models with lower dimensions of inputs, outputs and state spaces. However, due to the collinearity among the marketing inputs, as well as the collinearity among the marketing outputs, the number of dimensions of inputs, outputs and state spaces can be reduced during the identification of state space models.

Moreover, the objective function formatted can usually be nonlinear thus the guessing of the starting value is important before the implementation of the algorithms to estimate the parameter values.

In this research, we concentrate on the application of the ARX models for structural analysis and the method to estimate the parameters of state space models on brand equity dynamics. Due to unavailability of commercial data on brand performances and brand

investments in these years, for the research on the application of state space models, we propose procedures and methods to estimate these models, verify the effectiveness of these procedures and methods, but are not able to provide the implementation of state space models with real data. The empirical evidence is expected to be provided in future research.

The objectives of the research presented in this dissertation are as below:

1. Apply the ARX model to represent the dynamics of brand equity statistically to investigate brand equity dynamics and to obtain theoretical, managerial, and methodological implications.
2. Try to estimate the parameters of a state space model with one input using MLE method, together with the EM algorithm, develop appropriate method to guess the initial values of the parameters.
3. Estimate the parameters of a state space model with two inputs and nonlinear constraints, using MLE method, together with the EM algorithm.

The accomplishments of the objectives are respectively presented in the following three chapters. Among them, Chapter II is to estimate and apply ARX models. Chapter III is to estimate the parameters of a state space model with one input. The greatest efforts are on Chapter IV where more complicated procedures or approaches are developed to guess the initial values and estimate the parameter values. In addition, the reason why the MLE method and the EM algorithm is selected is explained in this chapter.

1.5 Contributions

Due to the differences between marketing research and engineering research, marketing research can not be implemented and evaluated all the same as an engineering

research. To bridge different research disciplines is challenging. However, as found by Robert et al. [42], brand management and tools with models are the research areas where marketing science has the largest impact on business decisions. The problem under investigation in this dissertation is interesting and of great value.

Although there are many challenges, as briefly introduced in Sect. 1.4.2. Several innovations and contributions are still accomplished by research.

1.5.1 Innovations and contributions

In Chapter II, through the application of the ARX models and the use of Interbrand data, a generic brand value structure analysis is performed. Brand outcome can be used for brand diagnostic purposes and to qualify brand long-term value and assess firm brand operation. From brand structural analysis, we propose to use brand intrinsic values and brand performance ratios to assess brands. Theoretically, we separate the long-term effects and short-term effects, and recommend intrinsic brand value and performance ratio in brand assessment. Managerially, the decision makers will be able to understand the structure of brand value, recognize the difference between the effects of investments on advertising and R&D, make smart decisions to optimize the returns from investment. Principle component analysis (PCR) and generalized difference are used in order to estimate the parameters of an ARX model.

In Chapter III, weighted least square (WLS) method is proposed to successfully to guess the initial parameter values of a simpler state space model which has only one time-varying exogenous input, one state and one output. The application of state space model with exogenous inputs may be not rare in engineering, models with time-varying exogenous inputs are rare in econometrics and time series analysis, especially when such

models are to be estimated. The estimation of a model is usually much more complicated than the application of a model. During the application of a model all the parameters of the model are already known, but by contrast, during the estimation of a model all the parameters of the model are unknown and must be obtained from inputs and outputs of the system. Only after the parameters are estimated, the application of the model is possible. In Chapter III, we take the first step to estimate state space models with time-varying exogenous inputs. This step is treated as the preparation for the work in Chapter IV.

In Chapter IV, during the estimation of a two-input state space model with nonlinear constraints, we design a procedure for the initial guessing of the parameter values and integrated optimization approaches during parameter estimation. As further introduced in Chapter IV, the guessing of initial parameter values is critical in the implementation of the EM algorithm. So the designed procedure is a great contribution to be referred by other researchers. In addition, the MCMC methods are usually used to estimate state space models with time-varying exogenous inputs and partly known parameters. We advocate integrating existing methods for mathematical optimization, instead of MCMC, in order to estimate the unknown parameters in a state space model. As a result, both the efficiency and the effectiveness of the estimation are improved though significantly reduce the computational complexity and the number of iterations of the EM algorithm. Through simulation, the procedure and the integration are proved to be promising in identifying a state space model with nonlinear constraints and exogenous inputs.

1.5.2 Challenges

The research covers the areas in signal processing, control engineering, econometric, time series analysis, mathematical optimization and marketing. It is

challenging to bridge between two domains: business and engineering. In this research, there are several challenges. The most significant challenges are caused by the differences between the research in marketing and engineering.

In engineering, certain fundamental concepts have unique definition and measurement. However, in marketing research, the definitions are not unique. As introduced, there are a few definitions about brand equity and brand value. In extant literature, there are more definitions about brand value than what has been introduced. In the mean while, there are different measures for brand equity, not limited to those mentioned in Sect. 1.2.

Consequently, there are different models to study brand equity in extant literature. These models are quite different with each other. As contrasted, in engineering, usually there is certain inheritance from an old model to a new model. Consequently, a researcher does not have to initiate a brand-new model and the advantages of a new model can be evaluated through the comparison between the old model and new model. In marketing research, it's rare to compare between different models.

Another way to evaluate a new model is to consider its application in industrial practice. The impact can be direct. An academic paper may be used to solve a practical problem. Usually, the impact is indirect. For example, contents of a paper may be incorporated into practitioners' tools, which then influence marketing decision making. However, as implicated by Roberts et al. [42], the impact of academic marketing research to marketing practice is not evaluated immediately, but it is in the following decades.

An issue relevant to marketing research is that commercial data are needed but firms are not mandated to provide certain data needed by the research. Some firms provide

the data for a certain period but not providing the data for another period. Thus, it is not guaranteed that the data needed for the research are complete. Moreover, commercial data, especially those in the recent period, are usually confidential and thus not available to researchers. This limits the booming of marketing research based on data analysis. Moreover, it is rare to forecast the future application in extant marketing literature. But, by contrast, data for engineering research can be obtained much easier; In some cases, simulation data can be used. It is common to provide recommendations to improve the performance in the current application.

1.6 Outline

In Chapter I, the motivation, background, objectives, and contributions of this research are presented.

In Chapter II, a generic brand value structure analysis to use brand outcome measures to generate further brand insights is presented. How the brand value structure analysis can use extant brand outcome measures to enhance brand assessment by analyzing fourteen top global brands is illustrated.

In Chapter III, approaches for both initial guessing and MLE of parameters of a state space model are proposed. In Chapter IV, a procedure is developed for initial guessing of values of the parameters, and mathematical optimization methods are integrated to obtain the estimates of the parameters which maximize the nonlinear objective function, when a parsimonious state space model for brand equity dynamics is employed. The simulation results are provided.

Chapter V concludes this dissertation and recommends future research directions.

Chapter II

ARX Models for Brand Value

Structure Analysis

2.1 Brand Structure Analysis

Brand equity measurement (i.e. assessing brand equity and value) is an important and basic research topic in the brand literature [43]. As the key to strategic planning and brand portfolio management, brand equity measurement, aligns with the tasks marketing executives performed and is of interest to researchers in both marketing and other business areas (e.g., accounting and finance respectively).

Taking different approaches, research has proposed many brand equity measures, including those based on the customer mind-set, product market (or company), and financial market [17] [43]. Customer mind-set measures assess the consumer-based *sources* of brand equity, such as awareness, attitudes, associations, attachments, and

loyalties. Both product market–based and financial market–based measures are brand performance *outcome*–based measures [17] that employ brand product market or financial market data to determine utility or dollar values of brands. Product market–based measures include the revenue premium measure [17], incremental choice probability measure [12], and consumer utility measure [36] [44] [45]. Examples of financial market–based measures are Interbrand’s brand value measures [46] and residual market value measures [14]. Shankar et.al [37] and Srinivasan et al. [47] provide brief summaries and comparisons of extant brand outcome measures.

Compared with measures based on customer mind-set, brand outcome measures provide a single, simple, more “complete,” objective measure of brand performance and are appealing to both marketing and non-marketing managers [17]. They conform with brand equity definitions—that is, brand equity is the value the brand adds to the product [48], or the marketing effects or outcomes that accrue to a product with a brand name compared with those that would accrue to the same product without the brand name [17]. However, brand outcome measures suffer from several disadvantages, which limit their usefulness in brand research and management.

First, because of their focus on outcomes rather than sources of brand equity, brand outcome measures lack diagnostic ability [17]. They can indicate strong or weak brand performance outcomes but cannot explain the reasons and sources. Scholars have continually called for research to link brand outcome measures to sources of brand equity [17] [43]. However, to our knowledge, only Srinivasan et al. [12] have tried to link brand incremental contribution, an outcome measure, to attribute- and non-attribute-based sources.

Second, outcome measures of brand equity do not differentiate brand long-term value from short-term performance. A brand, at least in concept, should be relatively long lasting and stable. However, driven by firm short-term strategy, such as brand investments and marketing action, brand performance can be highly volatile. For example, according to Interbrand's brand value, a financial market-based brand outcome measure, Sprite lost 16% (\$3,879 million to \$3,263 million) of its brand equity from 2008 to 2009 and increased brand equity by 77% (\$3,263 million to \$5,777 million) the next year (from 2009 to 2010). Louis Vuitton increased its brand equity by 144% (\$6,602 million to \$16,077 million) from 2004 to 2005. Sriram et al. (p. 71, Figure 2.1) [36] also show the volatility of brand outcome measures in their product market-based brand equity estimates. Without differentiating brand long-term value from short-term performance, brand outcome measures can mislead brand managers to be short-term focused. Ailawadi et.al [17] suggest that further research should quantify the long-term value of brand equity using brand outcome measures. In reviewing brand equity measures, Keller and Lehmann [43] suggest that research should separate brand impact from that of company market power and other possible determinants.

A potential way to address the disadvantages of brand outcome measures and make them more useful in brand research and management is to view brand as a market asset [49] - [51] and distinguish brand asset value from brand performance. According to the resource-based theories [52], firms have various capabilities to engage the resources and assets to produce competitive advantages; thus, it is necessary to distinguish firm capabilities from resources/assets. By separating firm brand capabilities (i.e. company market power by Keller and Lehmann 2006 [43] from brand as a resource, brand managers can gain better brand diagnostic and management insights.

In this chapter, we propose a generic brand value structure analysis so that brand outcome measures can be used for brand diagnostic purposes and to quantify brand long-term value and assess firm brand operation. Linking brand performance to brand equity sources on the basis of various customer brand responses, we decompose brand outcome measures into brand base performance, market inertia, and marketing-induced performance. Separating brand and brand operation in brand assessment, we identify and quantify brand intrinsic value (i.e. the long-term or persistent value a brand delivers without marketing actions) and the brand performance ratio (i.e. ratio of brand performance outcome to brand intrinsic value). With brand intrinsic value and performance ratio, we categorize brands into four types: (1) high intrinsic value, high performance ratio; (2) high intrinsic value, low performance ratio; (3) low intrinsic value, high performance ratio; and (4) low intrinsic value, low performance ratio. We suggest that brands with high intrinsic value have strong staying power, and brands with low intrinsic values and high performance ratios are operation dependent and more vulnerable to competition. Brands in different product markets should rely on brand loyalty and market inertia differently to build brand intrinsic value.

As we know, Interbrand's brand value can be used for academic research [18] [38] [53] - [57]. To illustrate the brand value structure analysis, we examine fourteen top global brands using Interbrand's brand values during the 1998–2012 period as the brand outcome measure. We show that the brand value structure analysis can reveal in-depth brand insights that brand outcome measures fail to capture. For example, brand intrinsic values reveal a different brand ranking from Interbrand's brand ranking. Brands like Sony should be ranked higher because Sony has a high intrinsic value, indicating its staying power, but

suffers from ineffective brand operations. Toyota, on the other hand, ranks lower because its outstanding brand performance is leveraged by effective brand operations. A comparison of Harley Davidson and HP shows that brands rely differently on brand base performance and market inertia to build intrinsic value. Harley Davidson relies more on brand base performance, i.e. a large portion of brand performance from brand loyal customers, and HP relies more on market popularity and network effects in branding. In addition to brand-level differences, our results also suggest market-level differences.

2.2 Theoretical Conceptualization

To understand and link brand performance outcome to brand equity sources, we draw on the brand loyalty literature and examine brand performance generated from different consumer–brand bonds or relationships. The brand loyalty literature suggests that not all loyal brand relationships are alike and that consumer–brand bonds or relationships exist with different strengths [58] - [60]. For example, using customer attitude and repeat patronage, Dick and Basu [21] suggest four types of loyalty relationships: loyalty (high attitude, high patronage), spurious loyalty or inertia (low attitude, high patronage), latent loyalty (high attitude, low patronage), and no loyalty (low attitude, low patronage). In general, brand loyalty research distinguishes inertia from loyalty on the basis of brand sensitivity [61] - [63]. Loyalty refers to strong brand sensitivity and commitment, and inertia of purchase involves low brand sensitivity and repeat purchase of a brand without a real motive [61].

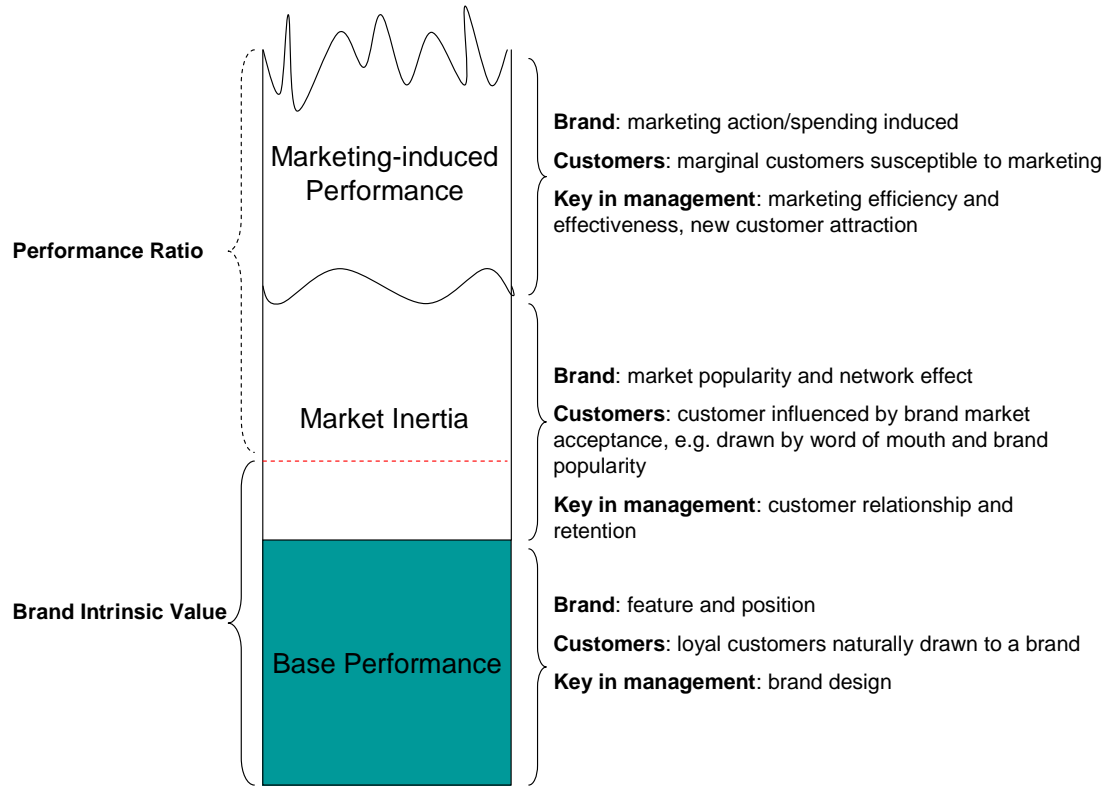


Figure 2.1: Decomposition of Brand Performance Outcomes

According to the loyal, inertial, and nonloyal relationships suggested in brand loyalty literature [61] [64] [65], at the aggregate level of consumer behavioral results, we consider three components of brand performance outcomes—namely, base performance, market inertia, and marketing-induced performance—and propose a brand value structure analysis. Figure 2.1 summarizes the three brand performance sources and the proposed key concepts in the brand value structure analysis (i.e. brand intrinsic value and performance ratio).

2.2.1 Brand value structure

Brand base performance refers to the brand performance generated through brand loyal customers. These customers have high brand commitment and are the core to a brand

because they are connected with and care about the brand [16]. Brand managers often refer to these customers as “true customers.” Since loyal customers like and appreciate a brand due to the fit between the brand’s features and positions and their needs, brand managers should focus on brand design, features, and positioning to increase brand base value. Whether the customers are hard-core loyalists who buy the brand all the time or split loyalists who are loyal to several brands [66], loyal customers generate stable brand performance outcomes. For mature brands, base performance can be modeled as a constant because the loyal customer group is relatively stable and purchases repeatedly with commitment.

Market inertia is the portion of brand performance related to brand performance of the previous period, and it comprises both the inertia of repeat purchase and new purchase from the brand network effect. With little brand commitment, brand inertia of repeat purchase is unstable [67] and may be caused and influenced by various factors, such as situational cues (e.g., convenience, familiarity), social influence (e.g., word of mouth) [21] [68] [69], and customer variety-seeking behaviour [70]. On an aggregate level, brand inertia of repeat purchase can be modeled as a percentage of brand performance of the previous periods and is considered a carryover effect [71] [72]. In addition, a brand’s market popularity and network effect may draw new customers. Some customers value a brand because their friends and/or colleagues use it, and others depend on word of mouth as an important information source. Brands become more valuable as they gain market acceptance and popularity, and customer relationship management and retention are important to these customers. A new purchase drawn by a brand’s market popularity and

network effect can also be evaluated by its relationship to brand performance of the previous period.

Marketing-induced performance refers to the short-term brand performance induced by brand marketing actions, spending, and operations, such as advertising and promotion. Marketing-induced performance is short-term oriented and highly fluctuating depending on firm short-term brand strategies and investment. The key activities in brand management are to improve efficiency and effectiveness of brand marketing spending and operation and to attract new customers.

2.2.2 Modeling brand value structure

According to the preceding analysis and understanding on brand base performance, market inertia, and marketing-induced performance, we model the brand value structure using an autoregressive model with exogenous inputs, considering lagged effects of market inertia and brand marketing actions:

$$BV_t = BV_{base} + \sum_{j=1}^p \alpha_j BV_{t-j} + \sum_{i=1}^n \sum_{h=0}^q \beta_{ih} MKT_{i,t-h} + e_t, \quad (2.1)$$

where BV_t is a brand outcome measure estimate at time t ; BV_{base} is brand base performance; $\alpha_j BV_{t-j}$ is the market inertia performance determined by brand performance of the previous periods ($t - p$ to $t - 1$) and the inertia ratios of each period (also called “carryover ratio” or “autoregressor”) [36] [26], with an expected range of

$0 \leq \alpha_j < 1$; $\sum_{i=1}^n \sum_{h=0}^q \beta_{ih} MKT_{i,t-h}$ is the marketing-induced performances caused by n

multiple brand marketing actions or spending (MKT_i) ($i = 1, \dots, n$), such as advertising and research and development (R&D), in current and previous periods ($t - q$ to t) where

h represents lagged time period; β_{ih} represents a firm's capability in transferring the type of brand spending to brand performance.

We model the brand value structure using Eq. 2.1 for the following reasons: First, Eq. 2.1 models the brand value structure with theoretical consistency and simplicity. That is, brand base performance is a constant, brand inertia pertains to brand performance of the previous periods, and brand marketing-induced performance is short-term oriented. Second, the reduced first-order autoregressive model of Eq. 2.1 ($h = 0$)

$$BV_t = BV_{base} + \alpha BV_{t-1} + [\beta_1, \beta_2, \dots, \beta_n] \begin{bmatrix} MKT_{1,t} \\ MKT_{2,t} \\ \dots \\ MKT_{n,t} \end{bmatrix} + e_t, \quad (2.2)$$

is of the same format of the model used by brand measurement research to validate proposed brand outcome measures and examine brand response to marketing-mix variables [17] [36] [73] - [75].

$$BV_t = C + \alpha BV_{t-1} + [\beta_1, \beta_2, \dots, \beta_n] \begin{bmatrix} MKT_{1,t} \\ MKT_{2,t} \\ \dots \\ MKT_{n,t} \end{bmatrix} + e_t, \quad (2.3)$$

where C is an unexplained constant. That is, the first-order autoregressive model is empirically accepted in literature to model brand performance. While pervious research ignored the constant item (i.e. C) in analysis and focused on brand inputs (i.e. MKT_i) and performance relationships (i.e. the short-term oriented marketing-induced brand performance), we provide a theoretical brand value structure analysis addressing the economic meanings of all parameters. Third, methodologically, this model is a generic

model that can be reduced to various models used in extant brand research. For example, when brand inertia is not considered (i.e. $\alpha_j = 0$) and advertising (ADV_t) and R&D (RD_t) spending are examined, Eq. 2.1 reduces to

$$BV_t = C + \beta_1 ADV_t + \beta_2 RD_t + e_t, \quad (2.4)$$

a commonly used regression model to examine brand drivers [76] [75]. Eq. 2.1 also covers the lag models used in extant literature [18] [14] [75] to examine the lagged effects of marketing inputs, such as advertising and R&D. Thus, the brand value structure described in Figure 2.1 and Eq. 2.1 is conceptually supported and methodologically simple and powerful in investigating brand dynamics.

2.2.3 Brand intrinsic value

By examining the brand value structure using Eq. 2.1, we quantify brand intrinsic value (i.e. the persistent value a brand delivers without any brand marketing action or spending) and brand performance ratio (i.e. brand performance in relation to brand intrinsic value). To identify brand intrinsic value, we follow Ailawaid et al. [17]'s suggestions and investigate the brand performance dynamics when a firm stops brand spending:

$$BV_t = BV_{base} + \sum_{j=1}^p \alpha_j BV_{t-j}. \quad (2.5)$$

For an easy illustration of the brand intrinsic value concept, consider the first-order model Eq. 2.2, where

$$BV_t = BV_{base} + \alpha BV_{t-1}, \quad (2.6)$$

$$BV_{t+n} = \frac{BV_{base}(1-\alpha^n)}{1-\alpha} + \alpha^n BV_t. \quad (2.7)$$

When $0 \leq \alpha < 1$ and a firm stops brand spending, brand performance outcomes can be broken down into two parts: The $\frac{BV_{base}(1-\alpha^n)}{1-\alpha}$ part is increasing at a decreasing rate, reflecting reinforced brand performance through customer purchase and repurchase experiences, and the $\alpha^n BV_t$ part is decaying, reflecting the lost brand stimulation and support from the firm. When $n \rightarrow \infty$, the brand performance reveals its intrinsic value—that is, the persistent portion a brand delivers without marketing inputs:

$$BV_{intrinsic} = \frac{BV_{base}}{1-\alpha}. \quad (2.8)$$

Similarly, for a general model Eq. 2.1,

$$BV_{intrinsic} = \frac{BV_{base}}{1 - \sum_{j=1}^p \alpha_j}, \quad (2.9)$$

where $\sum_{j=1}^p \alpha_j < 1$ is expected. As Eqs. 2.8 and 2.9 reveal, brand base value and brand inertia ratio determine brand intrinsic value.

Parenthetically, $0 < \sum_{j=1}^p \alpha_j < 1$ and $0 < \alpha < 1$ are expected range of the values in marketing practice. Technically, the values of $\sum_{j=1}^p \alpha_j$ and α can be any real numbers. In principle, when $|\sum_{j=1}^p \alpha_j| < 1$ or $|\alpha| < 1$, without marketing investments, the time series of brand value is stationary. That is, any stimulus to the brand value will decay and disappear. On the other hand, when $|\sum_{j=1}^p \alpha_j| \geq 1$ or $|\alpha| \geq 1$, even without marketing investments, the time series of brand value is evolving and exhibits random walk or excitement. Any change to the brand value will have permanent effects. That is, the

marketing investments will have long-lasting and permanent effects on brand value. In

practice, for most companies, $0 < \sum_{j=1}^p \alpha_j < 1$ and $0 < \alpha < 1$. In this chapter, we assume $0 <$

$\sum_{j=1}^p \alpha_j < 1$ and $0 < \alpha < 1$. The assumptions must be verified during empirical illustration.

In case this can not be verified to a certain firm, further investigation can be performed on this firm.

2.2.4 Brand performance ratio

Brand performance ratio evaluates brand performance in relation to brand intrinsic value. We define it as follows:

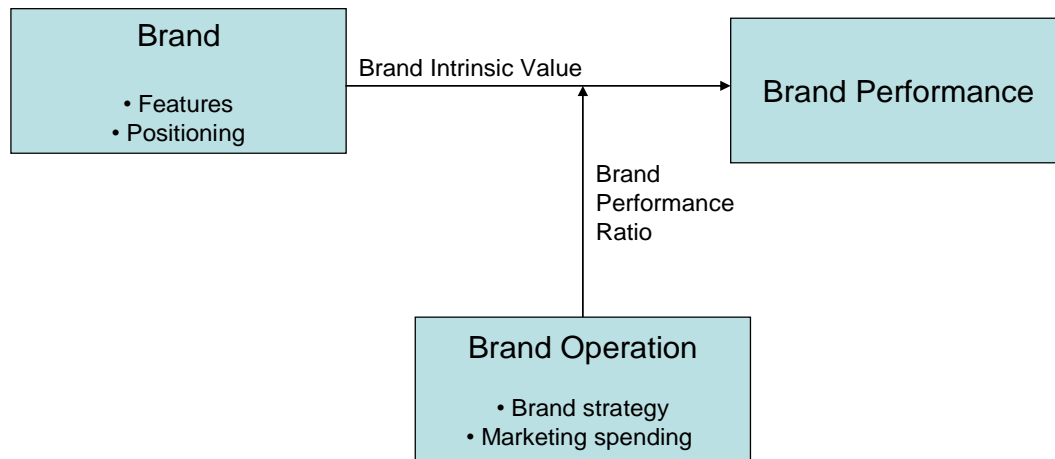


Figure 2.2: Brand and Brand Performance

$$r_{perf} = \frac{BV_t}{BV_{intrinsic}} . \quad (2.10)$$

Brand performance ratio reflects the overall effectiveness of brand strategies and firm brand value delivery capabilities. When $r_{perf} > 1$, firms' overall brand investments create positive brand value; conversely, $r_{perf} < 1$ indicates ineffective brand operation and investments, which result when brand performance is lower than its intrinsic value. In general, we expect that $r_{perf} > 1$. When $r_{perf} < 1$, immediate attention is needed to adjust brand short-term operation and investment strategies.

2.2.5 Implications from the brand value structure analysis

The brand value structure analysis links brand performance to brand equity sources on the basis of various consumer–brand relationships. By examining the three components of brand performance (i.e. base performance, market inertia, and marketing-induced performance) and a pair of brand evaluation measures (i.e. brand intrinsic value and performance ratio), we can draw several brand evaluation and management insights. First, brand intrinsic value and performance ratio are a pair of brand evaluation measures that are derived from brand outcome measures but effectively separate brand long-term value from short-term performance. Contrary to the volatile estimates of brand outcome measures, brand intrinsic value is stable during a period in which Eq. 2.1 holds unchanged. The brand performance ratio indicates the effectiveness of brand short-term operation and firm brand investment. As Figure 2.2 shows, brand intrinsic value and performance ratio provide a foundation to evaluate brand asset value and firm capability according to the resource-based theories [52].

Second, as Table 2.1 shows, depending on brand intrinsic value and performance ratio, we can categorize brand management into four types: (1) high intrinsic value, high performance ratio; (2) high intrinsic value, low performance ratio; (3) low intrinsic value, high performance ratio; and (4) low intrinsic value, low performance ratio. While apparently high-high is a desirable condition and low-low undesirable, in the other two combinations, high intrinsic value with low performance ratio is better than low intrinsic value with high performance ratio for brand management. Brands with high intrinsic values have a strong loyal customer base and staying power, and brands with low intrinsic values and high performance ratios are operation dependent and more vulnerable to competition.

Table 2.1: Four types of brand management scenarios

		Brand Intrinsic Value	
		High	Low
Brand Performance Ratio	High	Good brand, good brand operation: a desirable condition	Poor brand, good brand operation: Operational brands, vulnerable and low staying power
	Low	Good brand, poor operation: relatively high staying power	Poor brand, poor operation

Third, brand intrinsic value is determined by brand base value and market inertia ratio, not performance-driven brand spending. Thus, brand base value and inertia ratios are the most important elements in brand benchmarking and evaluation (i.e. building a loyal customer base and increasing brand market inertia ratio are important tasks). Two benchmarks can be performed. One is to benchmark brands with the best brand in the product market to identify gaps and perform what-if analysis to guide future brand strategy. The second is to compare the best brands in different product markets to identify the

effectiveness in using the loyal customer base and brand market inertia as brand-building tools in different product markets. Some markets may feature a high portion of loyal customers, while others may feature high market inertia.

2.3 Empirical Illustrations

To illustrate the brand value structure analysis, we analyze top global brands during the 1998–2012 period. Note that these empirical illustrations are not intended as a comprehensive empirical investigation of the performance of these brands; rather, they are employed to demonstrate the concepts we discussed. We use publicly available data to facilitate replication studies and theory verification.

2.3.1 Data

The main data sources available for use in this research are annual ranking of the best global brands released by Interbrand and Compustat provided by Standard and Poor (S&P). In case there are missing/wrong financial/marketing data for certain companies, especially in Compustat, annual 10-k reports or other government filings must be referred. Data from Yahoo! Finance and other resource are also considered for industrial classification (IC).

Interbrand is the first brand consulting company that meets the requirements of the ISO 10668 standard of monetary brand valuations. In addition, Interbrand started to release the list of top brands earlier than other organizations such as BrandZ, thus more brand value data are available from Interbrand. Interbrand estimates brand values using a discounted cash flow model and has been publishing annually its “Top 100 Global Brands” lists in *Financial World* during 1992-1997 (for brand values of 1991-1996) and then in

BusinessWeek since 1999. Interbrand's brand values have drawn much media and industry attention, and been used in marketing and accounting research to evaluate brand performance and examine brand response to marketing variables [24] [18] [75] [77] - [79]. In this chapter, we use brand values published in Interbrand's annual "Top 100 Global Brands" lists as the brand outcome measure in our illustration.

We limit our analysis to the 1998-2012 period to avoid potential problems of the unpublished 1997 brand values by Interbrand and valuation changes between Interbrand's publications in *Financial World* and *BusinessWeek*. We select only corporate brands, i.e. brands bearing the corporate names and accounting for the major part of firm revenues, to avoid potential problems when matching with firm brand investment data [78].

We use advertising and R&D expenses as two types of brand spending in the illustration and retrieve advertising and R&D expenditure data from Compustat. Compustat is a database of financial, statistical and market information on active and inactive global companies throughout the world. In Compustat, because data items such as the advertising expenditures (XAD), the R&D expenditures (XRD), and staff expenditures (XLR or XSF) are available only in the North America Fundamentals Annual database, this database is used in this research. It is a database of U.S. and Canadian fundamental and market information on active and inactive publicly held companies. It provides more than 300 annual income statement and 100 quarterly income statement, balance sheet, statement of cash flows, and supplemental data items. It also contains information on aggregates, industry segments, banks, market prices, dividends, and earnings. XAD represents the cost of advertising media (i.e. radio, television, and periodicals) and promotional expenses.

Table 2.2: Descriptive statistics and correlations

Brands	Variables	M (in B\$)	SD	Brand Value (BV)	Advertising (ADV)	R&D (RD)
Accenture	BV	7.33	1.26	1		
	ADV	.08	.01	.41	1	
	RD	.35	.11	.89***	.30	1
Amazon	BV	7.98	6.27	1		
	ADV	.48	.57	1.00***	1	
	RD	1.06	1.26	.99***	1.00***	1
Canon	BV	9.93	1.78	1		
	ADV	.95	.22	.72***	1	
	RD	2.97	.81	.86***	.81***	1
Dell	BV	9.97	1.90	1		
	ADV	.64	.21	.02	1	
	RD	.57	.21	-.57**	.70***	1
Ford	BV	14.16	9.09	1		
	ADV	3.79	.93	-.62**	1	
	RD	6.76	1.08	.31	.07	1
Harley Davidson	BV	5.95	1.82	1		
	ADV	0.07	.01	-.62**	1	
	RD	0.15	.03	.30	.39	1
Hewlett- Packard	BV	22.33	3.57	1		
	ADV	1.18	.30	-.45	1	
	RD	3.2	.50	.05	.43	1
Honda	BV	17.25	1.58	1		
	ADV	2.37	.44	.62**	1	
	RD	4.95	1.08	.78***	.70**	1
Intel	BV	33.84	3.07	1		
	ADV	1.92	.30	.08	1	
	RD	5.57	1.83	.37	.23	1
Microsoft	BV	59.94	2.61	1		
	ADV	1.32	.28	-.06	1	
	RD	7.15	1.94	-.62**	.52*	1
Philips	BV	7.03	1.96	1		
	ADV	1.18	.14	.61**	1	
	RD	2.63	.49	-.80***	-.54*	1
Sony	BV	11.71	1.76	1		
	ADV	4.07	0.55	-.42	1	
	RD	4.70	0.64	-.63**	.81***	1
Toyota	BV	27.75	5.13	1		
	ADV	3.55	.61	.82***	1	
	RD	7.86	1.91	.79***	.77***	1
Yahoo	BV	4.94	.87	1		
	ADV	.16	.05	.30	1	
	RD	.62	.46	.12	.80***	1

*** p -value < .01. ** p -value < .05. * p -value < .10.

The brands under investigation must meet a few requirements. Firstly, a company's brand shall be the only brand or the dominant brand. Secondly, the brand shall be in Interbrand's top 100 ranking during the years from 2001 to 2010. Thirdly, this firm shall be in Compustat's North America Fundamentals Annual. Fourthly, the brands/companies selected shall have data for advertising expenditures and R&D expenditures during the years from 2001 to 2010. Since these data are optional in companies' forms 10-K and other filings with the U.S. Securities and Exchange Commission (SEC), we must further check if these data items retrieved from Compustat are available and sufficient. Those companies or brands that cannot meet the requirements must be excluded from further investigation.

In addition, most data items we needed and obtained from Compustat are based on fiscal years, mainly because the annual data are from sources such as companies' annual reports and 10-k forms. On the contrast, brand value data is based on calendar. There is a need for time matching for data from these two resources. In this research, we performed some adjustments to the data items when the FYR (fiscal yearend month of the data), which identifies the month in which a company ends its fiscal year, is not December. Take the A&SP expense as an example, if we have the FYEAR (fiscal year), and the FYR, then to synchronize the fiscal year data from Compustat data with the data from Interbrand, we can obtain the data in calendar following below equations:

$$ASPYEAR = ASPFYEAR \times (12 - FYR) / 12 + ASPFYEAR - 1 \times FYR / 12, \text{ if } FYR \leq 5, \quad (2.11)$$

$$ASPYEAR = ASPFYEAR \times FYR / 12 + ASPFYEAR + 1 \times (12 - FYR) / 12, \text{ if } FYR \geq 6. \quad (2.12)$$

Similar adjustments are performed to *RD*. The cumulative adjustment factor allows us to evaluate market data on a comparable basis.

Finally, we identify 14 brands that have 15 years of reported advertising, R&D and brand values for analysis. They are Accenture, Amazon, Canon, Dell, Ford, Harley Davidson, Hewlett-Packard, Honda, Intel, Microsoft, Philips, Sony, Toyota, and Yahoo. Table 2.2 reports data descriptive statistics and correlations. In some cases, the correlations among brand value, advertising and R&D expenditures are high.

2.3.2 Procedures

We employ the following first-order autoregressive model as the empirical model for testing:

$$BV_t = BV_{base} + \alpha BV_{t-1} + \beta_{adv,-1} ADV_{t-1} + \beta_{rd} RD_t + \beta_{rd,-1} RD_{t-1} + e_t \quad (2.13)$$

Multicollinearity is observed between explanatory variables in Eq. 2.13. One cause of the multicollinearity is the lag effect of *ADV* and *RD* respectively. The other cause is a certain degree of correlation between *ADV* and *RD*. The existence of multicollinearity makes it difficult to identify the exact effect of a certain explanatory variable.

We intend to reduce the multicollinearity. However, there are no minor variables to be deleted, no additional constraints to delete a certain variable indirectly. The model shall neither be transformed. All in all, traditional approaches to solve the multicollinearity issues are not applicable in the research. Incidentally, reshuffling the data will not solve the problem because the data will be applied in regression. Consequently, principal component regression (PCR) [80] [81], in which principal component analysis (PCA) is used to mitigate multicollinearity among explanatory variables, shall be applied.

PCA is a multivariate data analysis technique applicable to models with multicollinearity. The essence of PCA is to use an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. These principal components independently represent the effects of the explanatory variables. The number of principal components is less than or equal to the number of original explanatory variables. An important property of the principal components is that the first principal components has the largest possible variance and each succeeding principal component in turn has the highest variance possible under the constraint that it be orthogonal to the preceding components. Therefore, the first principal component accounts for as much of the variability in the data as possible. The issue of multicollinearity can be solved through the selection of only a certain number of principal components whose cumulative variance is large enough to represent the whole variance of the original data set. The dependent variables can be explained by these selected principal components through linear regression. Through the relationship between the principal components and original explanatory variables, the original model regarding the dependent variable and original explanatory variables can be found. In such a model found, the impact of an explanatory variable is separated from the impact of another explanatory variable because PCA can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. In the process of PCA, the dimensionality of the transformed data is reduced thus the issue of multicollinearity can be solved. In addition, in order to remove the possible autocorrelation of the residue series generated from PCR, generalized difference is applied as a step during the proceeding of PCR.

The steps of PCA usually are

- (1) Standardize the original data set because PCA is sensitive to the scaling of the variables;
- (2) Calculate the eigenvalues and eigenvectors of the covariance matrix of the explanatory variables;
- (3) Normalize dependent variable (ZBV in this research), denote them as Z_1, Z_2, \dots , in sequence, find the relationship between principal components and original variables;
- (4) Perform regression on ZBV to principal components, check the autocorrelation and heteroscedasticity of the model;
- (5) Implement generalized difference estimation if necessary till the model with normalized variables is finalized;
- (6) Obtain the linear regression model regarding original variables using the relationship between the normalized variables and original variables

In this empirical illustration, PCA generates around two principal components for each of 14 regressions. Because the number of principal components used in regressions is less than the original explanatory variables in Eq. 2.13, PCA reduces the sample size requirement for acceptable prediction. Our sample sizes of around 12, which are determined by the number of years investigated, satisfy the sample size requirement for models using two principal components [82]. Based on principal component regression results, we calculate the brand intrinsic value and brand performance ratio for each brand according to Eqs. 2.8 and 2.10.

2.3.3 Endogeneity

To test for endogeneity (i.e. reverse causality between firm advertising and R&D expenditures and brand values), we follow Sriram et al. [36] and use log-transformed lagged values of advertising and lagged R&D as instruments. We then perform the Hausman–Wu test [83] [84]. Using a significant level of $p\text{-value} < 0.05$, the tests reveal no violation of the assumed exogeneity of advertising and R&D in our regression model.

2.3.4 Results

In econometrics, conventionally, the coefficient of determination, R^2 (or adjusted R^2), is widely accepted measure of how well observed outcomes are replicated by the model. R^2 is a number that indicates the proportion of the variance in the dependent variable that is predictable from the explanatory variable(s). Usually, R^2 ranges between 0 and 1. The regression model is good when R^2 is near 1. On the other hand, the quality of regression model cannot be directly measured by the mean square error of e_t . For the same mean square error of e_t , if the signal levels are different, the quality of regression is different.

Table 2.3 reports the principal component regression results. The regression model of Eq. 2.13 achieved a good fit, with all adjusted R^2 values higher than 0.39 and eleven out of fourteen higher than 0.60. Most parameter estimations are significant with p-values less than 0.10 (The p-value is a number between 0 and 1 and interpreted in the following way: A small p-value, typically smaller than 0.05, indicates strong evidence against the null hypothesis, so the null hypothesis is rejected). The brand base performance estimations range from 1.61 (Amazon) to 39.18 (Microsoft), all in billion dollars. The reported brand inertia ratios range from 0.18 (Honda) to 0.78 (Sony).

Table 2.3: Principal component regression analysis results

Brands	BV_{base}	α	β_{Adv}	β_{RD}	$b_{\text{Adv, t-1}}$	$b_{\text{RD, t-1}}$	Adj. R^2
Accenture	3.16**	.37***	26.84**	-35.83**	3.01***	3.89***	.96
Amazon	1.61***	.28***	2.18***	3.31***	1.00***	1.61***	.98
Canon	2.87**	.19***	1.45***	1.38***	.49***	.40***	.85
Dell	5.43***	.67***	1.41**	.38	-2.38***	-3.92***	.87
Ford	9.83**	.24**	-1.95**	-2.05**	1.17**	.96**	.77
Harley Davidson	6.56***	.36***	-39.17***	-48.22***	21.52***	-2.52***	.87
Hewlett-Packard	3.43	.66***	-2.64**	-4.16**	1.39*	2.40**	.88
Honda	9.04***	.18**	.53**	.53**	.30**	.27**	.54
Intel	19.50**	.47**	.14	-3.55*	.45**	.38*	.63
Microsoft	39.18**	.34*	.60	2.06	-.28*	-.25*	.39
Philips	4.29**	.26***	2.28***	2.63***	-.91***	-.85***	.85
Sony	5.40	.78**	.26	.14	-.44*	-.60*	.69
Toyota	4.19*	.21**	1.49**	1.47**	.54**	.47**	.69
Yahoo	1.80*	.47*	4.89**	-.68	.02	-.12	.42

*** $p\text{-value} \leq .01$; ** $p\text{-value} \leq .05$; * $p\text{-value} \leq .10$.

Table 2.4: Brand analysis

Brands	Proposed Brand Measures		M_{BV} (Interbrand)	Brand Ranking based on		Brand Characteristics
	$BV_{intrinsic}$	r_{perf}		$BV_{intrinsic}$	M_{BV}	
Accenture	5.01	1.46	7.33	11	11	A good brand with effective brand operation
Amazon	2.24	3.57	7.98	14	10	Brand performance far exceeding brand intrinsic value, very effective brand operation
Canon	3.57	2.78	9.92	12	9	Brand performance far exceeding brand intrinsic value, very effective brand operation; low market inertia.
Dell	16.43	.61	9.96	4	8	Ineffective brand operation, high brand potential; high market inertia.
Ford	12.86	1.10	14.16	5	5	A good brand with effective brand operation; low market inertia.
Harley Davidson	10.29	.58	5.95	7	13	High brand intrinsic value, ineffective brand operation, high brand potential
Hewlett-Packard	10.08	2.21	22.33	8	4	Highest brand intrinsic value, ineffective brand value delivery; high market inertia.
Honda	10.96	1.57	17.25	6	6	High brand intrinsic value, effective brand operation; low market inertia.
Intel	37.05	.91	33.84	2	2	Highest brand intrinsic value caused by both high brand base value and high market inertia but ineffective brand operation, high brand potential.
Microsoft	59.77	1.00	59.94	1	1	Highest brand intrinsic value caused by both high brand base value and high market inertia.
Philips	5.82	1.21	7.03	9	12	A good brand with effective brand operation
Sony	24.80	.47	11.70	3	7	Very high brand intrinsic value, but ineffective brand operation, high brand potential and staying power; high market inertia.
Toyota	5.30	5.23	27.73	10	3	Brand performance far exceeding brand intrinsic value, very effective brand operation; low market inertia.
Yahoo	3.39	1.46	4.94	13	14	A good brand with effective brand operation

Table 2.4 provides brand analysis based on regression results, including estimates of the proposed brand intrinsic values ($BV_{intrinsic}$) and brand performance ratios (r_{perf}), brand rankings (in the 14 brands analyzed) based on $BV_{intrinsic}$ and average Interbrand brand values during 1998-2012 (MBV), respectively, and summaries of brand characteristics. The brand intrinsic values range from 2.24 (Amazon) to 59.77 (Microsoft), all in billion dollars, and the brand performance ratios range from 0.47 (Sony) to 5.23 (Toyota).

To illustrate several different brand performance scenarios, we plot the Interbrand's brand value estimates (BV_t), brand intrinsic values ($BV_{intrinsic}$), advertising and R&D spending of four brands, i.e. Dell, Microsoft, Philips, and Yahoo, in Figure 2.3. In the case of Dell, brand intrinsic value is higher than brand (performance outcome) value. Microsoft demonstrates that the brand performance outcome approximately equals the brand intrinsic value. Philips' brand performance improves significantly over years. Yahoo's brand performance is consistently higher than the brand intrinsic value. To illustrate the ability of the empirical model (Eq. 2.13) to represent the sample data, we also plot predicted brand value (BV_{Est}). The predicted brand values closely track the actual Interbrand brand values.

As our results show, brand value structure analysis reveals brand insights that brand outcome measures fail to capture. By extracting brand intrinsic value and brand performance ratios from brand outcome performance (i.e. Interbrand's brand value), we evaluate brand and firm brand operations separately. First, shown in Table 2.4, brand ranking based on brand intrinsic value estimations is different from that of Interbrand's brand values. For example, Toyota's ranking in the fourteen brands decreases from number 3 based on Interbrand's value to number 10 based on brand intrinsic value; on the other hand, Sony's ranking increases from number 7 to 3.

Second, firms have different market power and brand operation effectiveness to leverage brand performance. Some brands, i.e. Harley Davidson, Intel and Sony, show ineffective brand operation with brand performance ratios less than one and need to improve their advertising and R&D strategies to fully explore brand potentials. Consider Sony as an example of a reputable and well-established brand. It is recognized and respected by customers as reliable and of high quality and has a high staying power. However, the rise of numerous competitors from countries such as South Korea and China has caused severe competition and saturated consumer electronics markets in the past decade, leading to decreased product differentiation and ineffective, and sometimes duplicated, R&D efforts among competitors. Sony has been losing its leadership in key categories, and needs to be more adaptable to customer needs and market competitions [46]. Our brand value structure analysis indicates that Sony needs to revitalize its brand operation through better brand strategies and management to deliver its intrinsic value. This insight is in line with Sony's focus on comeback strategies in practice in these years.

On the other hand, some brands such as Amazon, Canon, HP and Toyota benefit from highly effective brand operation through high brand investments and positive market response to brand investment, achieving high brand performance ratios ($r_{perf} \geq 2$). Consider Toyota as an example. Toyota's brand value (\$27.73B on average) is higher than that of Honda (\$17.25B on average), its close competitor. However, our brand value structure analysis suggests that Toyota ranks lower than Honda in brand intrinsic values, but operates on a much higher brand performance ratio. Toyota's brand suffered significantly from the 2009-11 vehicle recalls [85], which were widely covered by medias and followed by consumers. Several competitors, such as General Motors, Ford, Chrysler and Hyundai, took advantages of the Toyota crisis to steal

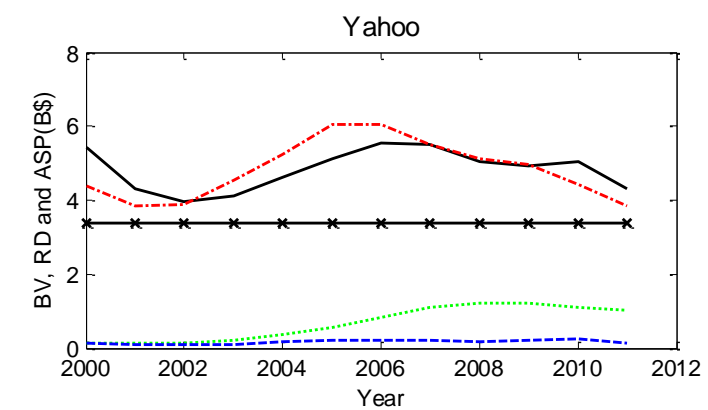
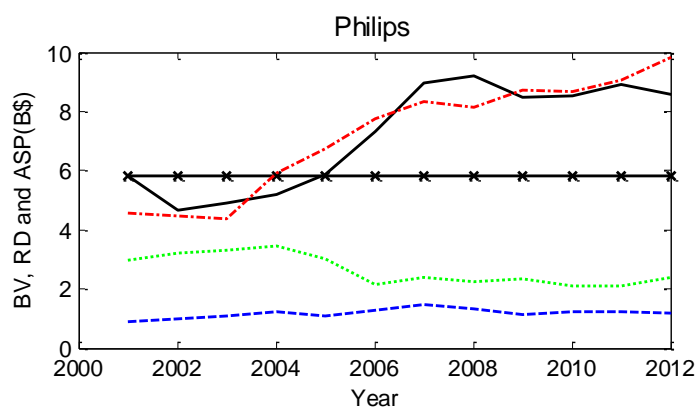
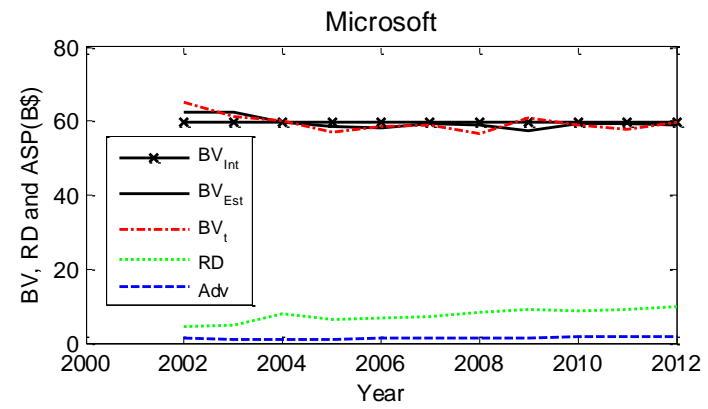
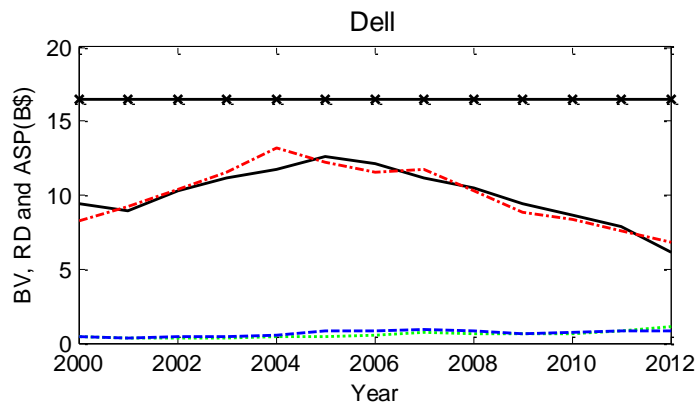


Figure 2.3: Brand Value, Brand Intrinsic Value, and Brand Marketing Spending

its loyal customers by offering cash rebates targeted toward Toyota owners. As a result, it is no surprising to find the lower brand intrinsic value of Toyota than Honda in our results. However, as the high brand performance ratio indicates, Toyota has strong market power. It continuously invests in brand advertising and R&D, higher than Honda's brand investments (see Table 2.2), and engages with customers [46]. The brand operation is successful, leading Toyota to quickly reclaim its global sales leadership position.

Third, brands rely differently on brand base performance and market inertia to build intrinsic value. For example, both Harley Davidson and HP have similar brand intrinsic values, i.e. \$10.29B and \$10.08B, respectively. However, a perusal of their brand value structure reveals important differences. Harley Davidson features a higher brand base performance than HP, and HP's market inertia ratio is about twice as high as Harley Davidson. This difference reflects the fact that comparing to HP, an IT supply and services brand, Harley Davidson relies more on its core loyal customers. Motorcycle buyers are a special group and Harley Davidson connects with its customers through a large and active brand community, which consists of clubs, events, and a museum. Harley Davidson supplies to many American police motorcycle fleets [86] and its logo licensing accounts for about 5% of total net revenue.

In addition to the brand-level difference, our results point to market-level differences. For example, the market inertia ratios of automobile brands, i.e. Ford ($\alpha=0.24$), Honda (0.18), and Toyota (0.21), are consistently lower than that of IT services brands, i.e. Accenture (0.37), Dell (0.67), HP (0.66) and Intel (0.47), in our brand sample. This suggests that comparing to the automobile market, word of mouth and brand popularity are even more important in the IT services market. Due to the high competition, quick product innovation, and short purchase cycle, customers of the IT services are highly sensitive to brand market acceptance. Situational

cues (e.g., familiarity) and social influence (e.g., word of mouth) [21] are often used in customer brand choice decisions.

2.4 General Discussion

Developing better measures of brand equity has been a research focus in the past two decades, and various brand outcome measures have been proposed [17] [47] [38]. These measures enable brand managers to track brand performance changes and investigate the effectiveness of brand marketing actions. However, because brand outcome measures focus on performance outcomes and do not link performance to brand equity sources, they lack explanatory power and diagnostic ability. In addition, because they do not distinguish brand long-term value from short-term performance, they can mislead brand managers to be short-term focused. Thus, determining how to use brand outcome measures to gain brand performance insights and quantify long-term value is an important research topic [17] [43].

This research proposes a generic brand value structure analysis to interpret estimates of existing brand outcome measures to gain further brand insights. The brand value structure analysis takes into account consumer–brand relationships as brand equity sources and decomposes brand performance measured by outcome measures into base performance, market inertia, and marketing-induced performance. Then, brand intrinsic value and brand performance ratio are defined and quantified to assess brand and brand operation, respectively.

2.4.1 Theoretical contributions

The brand value structure analysis and its derived measures contribute to literature on marketing research in the following ways: First, the analysis is one of few efforts in brand research to link brand performance outcomes to brand equity sources, enabling brand outcome

measures to be used for diagnostic purposes and become more useful in brand management. In contrast with Srinivasan et al. [47]'s work, which links brand incremental contribution to brand attribute- and non-attribute-based sources and requires additional survey data, our analyses borrow from brand loyalty research and use consumer–brand relationships as brand performance sources to interpret brand performance outcomes, limiting the need for further data collection. Brands that have similar performance outcomes may differ in the components of brand base performance, market inertia, and marketing-induced performance. A high portion of brand base performance is desirable because it represents a large loyal customer base.

Second, we quantify and distinguish long-term brand value from short-term performance. The brand value structure analysis provides economic explanation to an autoregressive model and suggests that brand intrinsic value is determined by brand base performance and market inertia ratio. This quantification is different from that of Ailawadi et al. [17], who estimate long-term brand value by the carryover coefficient (i.e. market inertia ratio) and discount rate and ignore brand base performance.

Third, we separate brand and brand operation in brand assessment and suggest that brand intrinsic value and performance ratio should be used to evaluate brand asset value and firm brand operation capability. To our knowledge, this is the first response in brand research to the suggestion of distinguishing firm capabilities from resources/assets in the resource-based theories [52] and the call to separate brand impact from that of company market power and other possible determinants [43]. Rather than focusing on brand performance outcomes, brand managers should categorize brands into four types on the basis of brand intrinsic value and performance ratio: (1) high intrinsic value, high performance ratio; (2) high intrinsic value, low performance ratio; (3) low intrinsic value, high performance ratio; and (4) low intrinsic value,

low performance ratio. Brands with high intrinsic values have strong staying power, and brands with low intrinsic values and high performance ratios are operation dependent and more vulnerable to competition. When $r_{perf} < 1$, a firm suffers from ineffective brand operation and fails to deliver brand intrinsic value.

Fourth, the generic brand value structure analysis can be applied to various brand outcome measures to gain brand insights. Ailawadi et al. [17] and Sriram et al. [38] use a similar model to Eq. 2.3 to validate their proposed revenue premium and consumer utility based brand equity measures and examine brand response to marketing-mix inputs. While we use the Interbrand's brand values to illustrate the brand value structure analysis, this generic brand value structure analysis can be applied to various brand outcome measures, which brand managers have on hand.

To demonstrate the brand value structure analysis and key measures proposed, we provide empirical illustrations by analyzing 14 top global brands in the recent 15 years and gain substantive results. By analyzing and comparing brand intrinsic values, two intrinsic value determinants (i.e. brand base value and market inertia), and performance ratio, we illustrate that (1) brand ranking based on the intrinsic value is different from that based on brand (performance outcome) value; As shown by Sony and Toyota, which brand value rankings differ significantly from brand intrinsic value rankings, focusing on brand performance outcomes without distinguishing brand from firm brand management can be misleading; (2) firms have different market power and brand operation effectiveness to leverage brand performance; in some cases (eg. Sony), firms fail to deliver brand intrinsic value and need short-term strategic adjustments; (3) brands rely differently on brand base performance and market inertia to build intrinsic value. For example, Harley Davidson and HP have similar

brand intrinsic values, but Harley Davidson relies more on brand base performance, i.e. a large portion of brand performance from brand loyal customers, and HP has a higher market inertia ratio, i.e. market popularity and network effects are prominent factors in branding; and (4) demonstrated by market inertia ratios of automobile and IT services brands, market-level differences may exist.

2.4.2 Managerial contributions

Brand managers continually monitor brand inputs and outputs and are equipped with various brand performance outcome measures. Traditionally, brand managers have used these measures to track brand performance changes, which often leads them to focus on short-term outcomes. For example, a reputable brand with lower market share may have lower value by brand outcome measures than a poor brand with higher market share. A brand may lose a significant share of its brand value evaluated by brand earnings when initiating a costly advertising campaign.

The brand structure analysis proposed herein enables brand managers to better use available brand outcome measures to understand brand performance structure (base performance, market inertia, and marketing-induced performance), features of consumer–brand relationships (loyal customers, less brand sensitive customers, nonloyal customers), and the importance of brand design and operation features (brand design, brand market effect, and marketing actions). With an in-depth understanding, what-if analysis can be carried out before brand planning and decisions are made. Brand managers can distinguish brand long-term value from short-term performance and evaluate overall operation success by examining brand intrinsic value and performance ratio. They can employ the brand structure analysis to

benchmark brands in the same product market to identify gaps or to compare the best brands in different product markets to understand market differences.

Our approach is simple and easy to use and does not require additional data collection efforts from managers. Brand managers already have the brand inputs and outputs data needed in this analysis. Although we use Interbrand's brand values as the brand outcome measure in our empirical illustrations, brand managers in practice can use other and likely better or more suitable brand outcome measures. We consider advertising and R&D expenditures in the empirical illustrations, but brand managers can consider other brand drivers. Although the true brand intrinsic value needs to be derived through consideration of a complete set of brand performance drivers, based on brand strategic analysis needs, focused analysis can be performed by considering a selective set of brand drivers. By assuming that the effect of other drivers remains stable, the focused analysis can reveal brand value structure information with respect to the brand drivers under study. Managers can also impose various conditions, such as budgeting constraints, to the model to suit their needs. Finally, we use annual data in the empirical illustration, but brand managers can use data with shorter intervals.

2.4.3 Methodological contributions

Due to the lag effects of *ADV* and *RD*, as well as the correlation between *ADV* and *RD*, Multicollinearity is observed between explanatory variables in Eq. 2.13. The existence of multicollinearity can greatly weaken the ARX model's ability to identify the exact effect of a certain explanatory variable.

Since conventional approaches to alleviate the multicollinearity issues are not applicable. PCR is performed to remove multicollinearity among explanatory variables through the selection of less number of principal components.

Moreover, generalized difference method is applied to mitigate the autocorrelation of the residue of the regression.

These methods are rarely used in extant study on the dynamics of brand equity.

2.4.4 Limitations and opportunities for further research

Many brand outcome measures have been proposed and validated in the brand research literature. However, no single measure satisfies all criteria of good brand measures [17]. Thus, comparing and choosing from available measures becomes important in both academic research and brand management practice. The presented brand value structure analysis provides a potential way to evaluate and compare brand outcome measures. Brand value structure analysis can be conducted using different brand outcome measures. The brand performance structure revealed through the analysis can be examined and compared with consumer–brand relationship survey data to validate brand outcome measures.

As the first step to decipher the brand value structure, we model a fixed brand intrinsic value (i.e. brand intrinsic value remains stable over an observation period) in Eq. 2.1. To use this method to achieve good estimations on brand value structure, we assume that brand intrinsic value stays relatively stable in the observation period. In practice, brand intrinsic value may change over time (e.g., in response to market changes and brand investments). Further research could model a changing brand intrinsic value, in response to market inputs and brand performance, using advanced modeling techniques.

For the purpose of demonstrating and explaining the proposed brand value structure analysis and the brand intrinsic value and performance ratio concepts, we analyze annual performance of 14 top global brands of the recent 15 years in our empirical illustration and gain substantive brand insights. This empirical illustration is not intended as a comprehensive

empirical study. The brand results can be further examined and validated in comprehensive empirical research with large sample size, short-interval brand performance data and more brand marketing input variables. For example, if a company has its monthly brand value data, which are measured and recorded by the company or provided by another company such as Interbrand, a close monitoring of the evolvement of brand equity and the effects of marketing activities is possible. Consequently, the effect of certain marketing campaign, such as recent Pepsi's Kendall Jenner commercial, can be estimated in time.

It shall be mentioned that the results of this research are applicable for the company itself only. Technically, each company or brand must implement ARX modeling with Interbrand BV data to find their own results. In the case that a company is in the same category with one of these 14 companies, the company shall only treat the results from the other company as references.

2.5 Chapter Summary

Performance outcome measures of brand equity have limited diagnostic ability and do not enable marketing managers to disentangle brand from brand operation. In this chapter, a generic brand value structure analysis is presented to use brand outcome measures to generate further brand insights. The brand value structure analysis (1) decomposes brand performance into base performance, market inertia, and marketing-induced performance and (2) defines and quantifies brand intrinsic value (i.e. the persistent value a brand delivers without marketing actions) and brand performance ratio (i.e. ratio of brand performance outcome to brand intrinsic value) to assess brand and brand operation, respectively. Brands with high intrinsic values have a strong loyal customer base and staying power while brands with low intrinsic

values and high performance ratios are operation dependent and more vulnerable to competition. Some brands fail to deliver intrinsic value because of ineffective brand strategy and investments. How the brand value structure analysis can use extant brand outcome measures to enhance brand assessment is illustrated by analyzing fourteen top global brands. In addition, it shall be emphasized that the application of PCR method and generalized difference are critical to guarantee the success of the research in this chapter.

Chapter III

A State Space Model and Its Maximum Likelihood Estimation

Form this chapter we work on the estimation of state space models which are expected to represent the dynamics of brand equity. As stated in Chapter I, the task to estimate a state space model is challenging. In order to fulfill the task step by step, in this chapter, we estimate a state space model which is relatively simpler than the model to be proposed in Chapter IV. The research in this chapter is to prepare for the more difficult research in the next chapter.

3.1 A State Space Model and MLE

One of the advantages of the state space models for linear dynamic systems is their ability to fit more parsimonious structures with fewer parameters to describe a multivariate

time series. As a result, the application of state space models is not limited to engineering practice.

For example, in market research, the dynamics of the brand equity of a firm can be represented by a state space model,

$$\begin{cases} x_t = \alpha x_{t-1} + \gamma u_t + e_t \\ y_t = x_t + w_t \end{cases}, \quad (3.1)$$

where x_t is the invisible brand equity, u_t is the investment, the exogenous input, y_t is the brand performance, the output, at a certain step t . The transition coefficient, α , and the input coefficient, γ , are parameters of interest. We assume that $|\alpha| < 1$ as explained in Chapter II. The observation noise, w_t , the process noise, e_t , and the initial state, x_0 , at each step $\{x_0, e_1, \dots, e_t, w_1 \dots w_t\}$ are all assumed to be mutually independent, where $w_t \sim N(0, \sigma_w^2)$, $e_t \sim N(0, \sigma_e^2)$, and $x_0 \sim N(\mu_0, \sigma_0^2)$.

This state space mode has an exogenous input. For marketing research, there may be more than one input. However, in some special market, there may only be one type of investment. Moreover, in the cases when there are two or more types of investments, it is common that the time series of these types of investments are collinear. After applying PCA, only one principle component is kept. As a result, a state space model with only one exogenous input is appropriate to investigate the BE dynamics of a firm with certain type of investments.

It shall be noticed that the state space models used in this chapter, as well as Chapter IV, are different from typical discrete-time state space models in the time argument for the input which should normally be the same as the state in the right-hand side of the equation ($t-1$ instead of t) [41] (pp. 93-104). The typical discrete-time state space models are evolved

from continuous-time models for physical systems. However, in areas other than engineering, untypical discrete-time state space models are also used, because those models are not constrained by the requirements physically. For instance, in an economic system, a current input may affect a current state because the investigation is usually based on periods of a week, a month, and even a year. Therefore, it is reasonable to use untypical state space model. In extant literature, both Van Heerde et al. [97] and Ataman et al. [98] used such untypical models.

Maximum likelihood estimation (MLE) [87] is used to obtain the time-invariant parameters of a state space model with known time series about input, u_t , and the output, y_t . Because the Kalman filter for the state space model operates recursively on streams of noisy input data to produce a statistically optimal estimate of the underlying system state, the mathematical optimization problem is a problem of nonlinear global optimization, where the objective function has a large number of local minima and maxima. Classical methods to find maximum likelihood estimates, such as gradient descent, conjugate gradient or variations of the Gauss–Newton method, which typically require the evaluation of first and/or second derivatives of the likelihood function, are often trapped at one of many local optima, especially when the sample size is large. Moreover, symbolic (analytical) methods are frequently not applicable, and the use of numerical solution strategies often leads to very complex challenges. Consequently, expectation-maximization (EM) algorithm [88], an iterative method, is used to complete MLE of parameters in state space models.

Typical approaches for the implementation of the EM algorithm in the MLE of linear state space model are delivered by Ghahramani and Hinton [89], Shumway and Stoffer [90] and Holmes [91]. However, estimation approaches for state space models with time-varying exogenous inputs are not provided in extant literature. Although in [91], [92] and [94], the inputs in the state space mode are either constant or not affecting the hidden variables.

Our research in this chapter is to find appropriate methods for the supervised problem: using the EM algorithm for the MLE of those state space models which have exogenous inputs. Considering the complexity of the issue of state space model estimation, the research is a preparatory attempt to estimate more complicated state space models such as the model in Chapter IV.

In this chapter, we use the model represented by Eq. 3.1 as an example of the models with exogenous inputs. Our target is to estimate the parameters, α , γ , σ_e , σ_w , μ_0 and σ_0 . Our work is carried out mainly in two phases: initial guessing and EM estimating with approaches different from ones in extant literature. Moreover, if the sample size is large, we use the asymptotic variances of the estimated parameters to check the accuracy of the estimation. If the sample size is small, we introduced bootstrapping to examine of the distribution of the estimated parameters.

3.2 Initial Guessing

The initial guessing is performed through two steps. Firstly, the system parameter α and the variances σ_e , σ_w and σ_0 are guessed using autocovariance of the observations.

Then the initial state mean μ_0 and control parameter γ are guessed using weighted linear square (WLS).

3.2.1 Initial guessing of the system parameter and the variances

Denote $\eta_t = \sum_{i=0}^{\infty} \alpha^i e_{t-i}$, from Eq. 3.1 we have

$$x_t = \frac{\gamma u_t}{1-\alpha} + \eta_t . \quad (3.2)$$

For $h = 0, 1, 2 \dots$, we can obtain both the covariance (when $h = 0$) and autocovariance (when $h > 0$) of η_t as

$$\gamma_{\eta}^{(h)} = \frac{\sigma_e^2 \alpha^h}{1-\alpha^2}, \quad (3.3)$$

while it is obvious that

$$\sigma_0^2 = \frac{\sigma_e^2}{1-\alpha^2}, \quad (3.4)$$

which means that we do not have to estimate σ_e and σ_0 separately.

Moreover, we have the variance of y_t

$$\gamma_y^{(0)} = \sigma_0^2 + \sigma_w^2, \quad (3.5)$$

and the covariance of y_t when $h = 1, 2, 3 \dots$,

$$\gamma_y^{(h)} = \gamma_{\eta}^{(h)}. \quad (3.6)$$

Hence we can obtain the initial guess for α , σ_e and σ_w

$$\begin{cases} \alpha = \frac{\gamma_y^{(2)}}{\gamma_y^{(1)}} \\ \sigma_e^2 = \frac{(1-\alpha^2)\gamma_y^{(1)}}{\alpha} \\ \sigma_w^2 = \gamma_y^{(0)} - \frac{\sigma_e^2}{1-\alpha^2} \end{cases} . \quad (3.7)$$

where $\gamma_y(1)$, instead of $\gamma_y(0)$, is used to calculate α and σ_e^2 because of the deference in the calculation of $\gamma_y(0)$ (Eq. 3.5) and $\gamma_y(0)$ (Eq. 3.6).

3.2.2 Initial state mean and control parameter

Denoting

$$z_t = \sum_{i=1}^t \alpha^{t-i} u_i, \quad (3.8)$$

and

$$\zeta_t = \sum_{i=1}^t \alpha^{t-i} e_i + w_t, \quad (3.9)$$

we have,

$$y_t = \alpha^t \mu_0 + \gamma z_t + \zeta_t, \quad (3.10)$$

where $\zeta_t \sim N(0, \sigma_\zeta^2)$ and

$$\sigma_\zeta^2 = \frac{1-\alpha^{2t}}{1-\alpha^2} \sigma_e^2 + \sigma_w^2. \quad (3.11)$$

In order to estimate μ_0 and γ , we perform linear regression between y_t , as dependent variable, and α^t and z_t , as independent variables, using T sample of u_t and y_t . Since ζ_t is heteroscedastical, we apply WLS, which finds its optimum when the weighted sum, S , of squared residuals is minimized where

$$S = \sum_{t=1}^T \frac{(y_t - \alpha^t \mu_0 - \gamma z_t)^2}{\frac{1-\alpha^{2t}}{1-\alpha^2} \sigma_e^2 + \sigma_w^2}. \quad (3.12)$$

Denote $v_t = (1 - \alpha^{2t}) \sigma_e^2 + (1 - \alpha^2) \sigma_w^2$, we solve the gradient equation (regarding μ_0 and γ respectively) for the sum of squares

$$\begin{cases} \sum_{t=1}^T \frac{\alpha^t (y_t - \alpha^t \mu_0 - \gamma z_t) (1 - \alpha^2)}{v_t} = 0 \\ \sum_{t=1}^T \frac{z_t (y_t - \alpha^t \mu_0 - \gamma z_t) (1 - \alpha^2)}{v_t} = 0 \end{cases}. \quad (3.13)$$

Therefore, we will have initial guess about μ_0 and γ as below:

$$\begin{cases} \gamma = \frac{\sum_{t=1}^T \frac{\alpha^2}{v_t} \sum_{t=1}^T \frac{z_t y_t}{v_t} - \sum_{t=1}^T \frac{\alpha^t y_t}{v_t} \sum_{t=1}^T \frac{\alpha^t z_t}{v_t}}{\sum_{t=1}^T \frac{\alpha^2}{v_t} \sum_{t=1}^T \frac{z_t^2}{v_t} - \sum_{t=1}^T \frac{\alpha^t z_t}{v_t} \sum_{t=1}^T \frac{\alpha^t z_t}{v_t}} \\ \mu_0 = \frac{\sum_{t=1}^T \frac{\alpha^t y_t}{v_t} \sum_{t=1}^T \frac{z_t^2}{v_t} - \sum_{t=1}^T \frac{\alpha^t z_t}{v_t} \sum_{t=1}^T \frac{y_t z_t}{v_t}}{\sum_{t=1}^T \frac{\alpha^2}{v_t} \sum_{t=1}^T \frac{z_t^2}{v_t} - \sum_{t=1}^T \frac{\alpha^t z_t}{v_t} \sum_{t=1}^T \frac{\alpha^t z_t}{v_t}} \end{cases} \quad (3.14)$$

3.3 Estimation Using the EM Algorithm

We can write the conditional density for the states and outputs,

$$P(x_t | x_{t-1}) = \frac{1}{\sigma_e \sqrt{2\pi}} \exp \left[-\frac{(x_t - \alpha x_{t-1} - \gamma u_t)^2}{2\sigma_e^2} \right], \quad (3.15)$$

$$P(y_t | x_t) = \frac{1}{\sigma_w \sqrt{2\pi}} \exp \left[-\frac{(y_t - x_t)^2}{2\sigma_w^2} \right]. \quad (3.16)$$

Assuming a Gaussian initial state density

$$P(x_0) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[-\frac{(x_0 - \mu_0)^2}{2\sigma_0^2} \right], \quad (3.17)$$

By the Markov property implicit in this model, we calculate the joint probability, not the partial probability used by Shumway and Stoffer [90], regarding all T samples of x_t and y_t , denoted as $\{x\}$ and $\{y\}$ respectively:

$$P(\{x\}, \{y\}) = P(x_0) \prod_{t=1}^T P(x_t | x_{t-1}) \prod_{t=1}^T P(y_t | x_t) \quad (3.18)$$

We denote the joint log probability as

$$\Omega = \log P(\{x\}, \{y\}). \quad (3.19)$$

According to Eq. 3.4, we only need to estimate the parameter set, $\Psi = \{\alpha, \gamma, \sigma_e, \sigma_w, \mu_0\}$, through maximizing the objective function:

$$\Omega(\alpha, \gamma, \sigma_e, \sigma_w, \mu_0) = -\frac{(1-\alpha^2)(x_0-\mu_0)^2}{2\sigma_e^2} - \frac{1}{2} \ln \frac{\sigma_e^2}{1-\alpha^2} - \sum_{t=1}^T \frac{(x_t - \alpha x_{t-1} - \gamma u_t)^2}{2\sigma_e^2} - \frac{1}{2} T \ln \sigma_e^2 - \sum_{t=1}^T \frac{(y_t - x_t)^2}{2\sigma_w^2} - \frac{1}{2} T \ln \sigma_w^2 - \frac{2T+1}{2} \ln(2\pi). \quad (3.20)$$

3.3.1. The EM algorithm

Since the objective function given above depends on the unobserved data series, x_t , $t = 1, 2, \dots, T$, we consider applying the EM algorithm conditionally with respect to the observed output series y_1, y_2, \dots, y_T . Different from the EM algorithm for linear dynamic systems proposed by Shumway and Stoffer [90], the objective function expressed in Eq. 3.20 has the inputs thus the input coefficient must be estimated.

In the E Step of the EM algorithm, the parameters are assumed known, the hidden states and their variance are estimated over all the samples, and then the likelihood function constructed from joint probability is calculated.

For the M step in the EM algorithm, we must find the parameter set $\hat{\Psi}(k) = \{\alpha(k), \gamma(k), \sigma_e(k), \sigma_w(k), \mu_0(k)\}$ during the k th counts of the recursions by maximizing the conditional expectation, or the objective function as in Eq. 3.20.

The overall procedure for the estimation is as below:

- (i) Initialize the procedure by selecting the guessed values as starting values for the parameters.

On iteration k , ($k = 1, 2, \dots$)

- (ii) Compute the log-likelihood (optional).
- (iii) Use the parameters, obtain the smoothed values of the hidden states and their correlations, for $t = 1, 2, \dots, T$.

- (iv) Use the smoothed values to calculate the updated parameters.
- (v) Repeat steps (ii) – (iv) to convergence.

The EM algorithm always increase the likelihood and is guaranteed convergence to a stationary point for an exponential family [93].

We mainly perform two sub-steps in the E step of the EM algorithm: Kalman filtering and Kalman smoothing.

3.3.2 Kalman filtering and smoothing

Assuming that we already know the parameter set $\{\alpha, \gamma, \sigma_e, \sigma_w, \mu_0, \sigma_0\}$ ($x_0 \sim N(\mu_0, \sigma_0^2)$), and the observations y_t and inputs u_t , we have the estimation of the hidden states, as well as the variances estimated based on the observations for the period 1 to t .

$$x_{t|t-1} = \alpha x_{t-1|t-1} + \gamma u_t, \quad (3.21a)$$

$$v_{t|t-1} = \alpha^2 v_{t-1|t-1} + \sigma_e^2, \quad (3.21b)$$

$$\tilde{y}_t = y_t - x_{t|t-1}, \quad (3.21c)$$

$$s_t = v_{t|t-1} + v_w, \quad (3.21d)$$

$$k_t = v_{t|t-1} s_t^{-1}, \quad (3.21e)$$

$$x_t^t = x_{t|t-1} + k_t \tilde{y}_t, \quad (3.21f)$$

$$v_t^t = (1 - k_t) v_{t|t-1}, \quad (3.21g)$$

where $x_{0|0} = \mu_0$ and $v_{0|0} = \sigma_0^2$.

According to Shumway and Stoffer [90], to compute $E[x_t|\{y, u\}] \equiv x_t^T$ and the correlation matrices $p_t \equiv v_t^T + x_t^T (x_t^T)'$ one performs a set of backward recursions using

$$j_{t-1} = \alpha \frac{v_{t-1|t-1}}{v_{t|t-1}}, \quad (3.22)$$

$$x_{t-1}^T = x_{t-1|t-1} + j_{t-1}(x_t^T - \alpha x_{t-1|t-1} - \gamma u_t), \quad (3.23)$$

$$v_{t-1}^T = v_{t-1|t-1} + j_{t-1}(v_t^T - v_{t|t-1})j'_{t-1}, \quad (3.24)$$

where $x_T^T = x_{T|T}$ and $v_T^T = v_{T|T}$. We also have $p_{t,t-1} \equiv v_{t,t-1}^T + x_t^T(x_{t-1}^T)'$, where $v_{t,t-1}^T$

can be obtained through the backward recursions

$$v_{t-1,t-2}^T = v_{t-1|t-1}j'_{t-2} + j_{t-1}(v_{t,t-1}^T - \alpha v_{t-1|t-1})j'_{t-2}, \quad (3.25)$$

which is initialized using $v_{T,T-1}^T = \alpha(1 - k_T)v_{T-1}^{T-1}$.

Note that the state estimate, x_t^T , differs from that computed by the Kalman filter in that it is the minimum mean square error smoothed estimator of x_t based on all observed data (and input data), i.e. it depends on past and future observations; the Kalman filter estimates $E[x_t|\{y\}_1^t]$ are the estimators based on past and future observations.

3.3.3 Expected log-likelihood formulation

After we have got the expected values for x_0 and x_t as $x_0^T \equiv E[x_0|\{y, u\}]$ and $x_t^T \equiv E[x_t|\{y, u\}]$ respectively, we can calculate the expectation of the log-likelihood

$$E(\Omega) = E[\ln P(\{x\}, \{y\})]. \quad (3.26)$$

Applying Eq. 3.1, $E(\Omega)$ becomes

$$E(\Omega) = -\frac{1}{2}E(\Omega_0) - \frac{1}{2}E(\Omega_1) - \frac{1}{2}E(\Omega_2) - \frac{2T+1}{2}\ln(2\pi), \quad (3.27)$$

where

$$\begin{aligned} \Omega_0 &= \frac{(x_0 - \mu_0)^2}{v_0} + \ln v_0, \\ \Omega_1 &= \sum_{t=1}^T \frac{(x_t - \alpha x_{t-1})^2}{v_e} + T \ln v_e, \\ \Omega_2 &= \sum_{t=1}^T \frac{(y_t - x_t)^2}{v_w} + T \ln v_w, \end{aligned}$$

Denote

$$\begin{cases} P_{t-1}^T = \sum_{t=1}^T E[x_{t-1}x'_{t-1}|\{y\}] \\ P_t^T = \sum_{t=1}^T E[x_t x'_t|\{y\}] \\ P_{t,t-1}^T = \sum_{t=1}^T E[x_t x'_{t-1}|\{y\}] \end{cases},$$

$$\begin{cases} M_{t-1}^T = \sum_{t=1}^T E(x_{t-1}u_t) \\ M_t^T = \sum_{t=1}^T E(x_t u_t) \\ U_t^T = \sum_{t=1}^T u_t^2 \end{cases},$$

and

$$\begin{cases} W_{t-1}^T = \sum_{t=1}^T E(x_{t-1}y_t) \\ W_t^T = \sum_{t=1}^T E(x_t y_t) \\ Y_t^T = \sum_{t=1}^T y_t^2 \end{cases},$$

we have

$$E(\Omega_0) = \frac{1-\alpha^2}{\sigma_e^2} [v_0^T + (x_0^T - \mu_0)^2] + \ln \frac{\sigma_e^2}{1-\alpha^2}, \quad (3.27a)$$

$$E(\Omega_1) = \sigma_e^{-2} (P_t^T + \alpha^2 P_{t-1}^T + \gamma^2 U_t^T - 2\alpha P_{t,t-1}^T - 2\gamma M_t^T + 2\alpha\gamma M_{t-1}^T) + T \ln \sigma_e^2, \quad (3.27b)$$

$$E(\Omega_2) = \sigma_w^{-2} (Y_t^T + P_t^T - 2W_t^T) + T \ln \sigma_w^2. \quad (3.27c)$$

3.3.4 Estimation of the parameters

We use the first order conditions on partial derivatives of $E(\Omega)$ to individual parameters to obtain the gradient and then the values of individual parameters. This method is not the multivariate regression approach used by Shumway and Stoffer [90]. The parameters are chosen such that the objective function is maximized, i.e., the gradients, i.e. the first order partial derivatives, are all zero., i.e.

$$\begin{cases} \frac{\delta E(\Omega)}{\delta \mu_0} = 0 \\ \frac{\delta E(\Omega)}{\delta \alpha} = 0 \\ \frac{\delta E(\Omega)}{\delta \gamma} = 0 \\ \frac{\delta E(\Omega)}{\delta \sigma_e} = 0 \\ \frac{\delta E(\Omega)}{\delta \sigma_w} = 0 \end{cases}.$$

Considering Eq. 3.27, we have,

$$\left\{ \begin{array}{l} \frac{\delta E(\Omega_0)}{\delta \mu_0} = \frac{2(x_0^T - \mu_0)}{\sigma_0^2} = 0 \\ \frac{\delta E(\Omega_1)}{\delta \alpha} = 2\sigma_e^{-2}(\alpha P_{t-1}^T - P_{t,t-1}^T + \gamma M_{t-1}^T) = 0 \\ \frac{\delta E(\Omega_1)}{\delta \gamma} = 2\sigma_e^{-2}(\gamma U_t^T - M_t^T + \alpha M_{t-1}^T) = 0 \\ \frac{\delta E(\Omega_1)}{\delta \sigma_e} + \frac{\delta E(\Omega_0)}{\delta \sigma_e} \\ \quad = -2 \frac{P_t^T + \alpha^2 P_{t-1}^T + \gamma^2 U_t^T - 2\alpha P_{t,t-1}^T - 2\gamma M_t^T + 2\alpha\gamma M_{t-1}^T + (1-\alpha^2)[v_0^T + (x_0^T - \mu_0)^2]}{\sigma_e^3} \\ \quad \quad \quad + 2 \frac{T+1-\alpha^2}{\sigma_e} = 0 \\ \frac{\delta E(\Omega_2)}{\delta \sigma_w} = -2\sigma_w^{-3}(Y_t^T + P_t^T - 2W_t^T) + 2T\sigma_w^{-1} = 0 \end{array} \right. .$$

The estimate of α , γ , σ_e^2 , σ_w^2 and μ_0 are from below five equations

$$\alpha(1 - \alpha^2)v_0^T - \alpha\sigma_e^2 - \alpha(1 - \alpha^2)P_{t-1}^T + (1 - \alpha^2)P_{t,t-1}^T - \gamma(1 - \alpha^2)M_{t-1}^T = 0, \quad (3.28a)$$

$$\gamma U_t^T - M_t^T + \alpha M_{t-1}^T = 0, \quad (3.28b)$$

$$(1 - \alpha^2)v_0^T - (1 + T)\sigma_e^2 + P_t^T + \alpha^2 P_{t-1}^T + \gamma^2 U_t^T - 2\alpha P_{t,t-1}^T - 2\gamma M_t^T + 2\alpha\gamma M_{t-1}^T = 0, \quad (3.28c)$$

$$\sigma_w^2 = \frac{1}{T}(Y_t^T + P_t^T - 2W_t^T), \quad (3.28d)$$

$$\mu_0 = x_0^T. \quad (3.28e)$$

Moreover, we use the second order partial derivatives of $E(\Omega)$ to calculate the information matrix. Most of them are zero except those listed below:

$$\frac{\partial^2 E(\Omega)}{\partial \alpha^2} = \frac{v_0^T + (x_0^T - \mu_0)^2 - P_{t-1}^T}{\sigma_e^2} - \frac{1 + \alpha^2}{(1 - \alpha^2)^2}, \quad (3.29)$$

$$\frac{\partial^2 E(\Omega)}{\partial \gamma \partial \alpha} = \frac{\partial^2 E(\Omega)}{\partial \alpha \partial \gamma} = -\frac{M_{t-1}^T}{\sigma_e^2}, \quad (3.30)$$

$$\frac{\partial^2 E(\Omega)}{\partial \sigma_e \partial \alpha} = \frac{\partial^2 E(\Omega)}{\partial \alpha \partial \sigma_e} = \frac{2(\alpha P_{t-1}^T - P_{t,t-1}^T + \gamma M_{t-1}^T) - 2\alpha[v_0^T + (x_0^T - \mu_0)^2]}{\sigma_e^3}, \quad (3.31)$$

$$\frac{\partial^2 E(\Omega)}{\partial \gamma^2} = -\frac{U_t^T}{\sigma_e^2}, \quad (3.32)$$

$$\frac{\partial^2 E(\Omega)}{\partial \sigma_e \partial \gamma} = \frac{\partial^2 E(\Omega)}{\partial \gamma \partial \sigma_e} = \frac{2(\gamma U_t^T - M_t^T + \alpha M_{t-1}^T)}{\sigma_e^3}, \quad (3.33)$$

$$\frac{\partial^2 E(\Omega)}{\partial \sigma_e^2} = \frac{T+1}{\sigma_e^2} - \frac{3(1-\alpha^2)[v_0^T + (x_0^T - \mu_0)^2]}{\sigma_e^4} - \frac{3(P_t^T + \alpha^2 P_{t-1}^T + \gamma^2 U_t^T - 2\alpha P_{t,t-1}^T - 2\gamma M_t^T + 2\alpha\gamma M_{t-1}^T)}{\sigma_e^4}, \quad (3.34)$$

$$\frac{\partial^2 E(\Omega)}{\partial \sigma_w^2} = \frac{T}{\sigma_w^2} - \frac{3(Y_t^T + P_t^T - 2W_t^T)}{\sigma_w^4}, \quad (3.35)$$

$$\frac{\partial^2 E(\Omega)}{\partial \mu_0^2} = -\frac{1 - \alpha^2}{\sigma_e^2}. \quad (3.36)$$

In addition, when all the first order derivatives equal zero, the likelihood is optimized as,

$$E(\Omega) = -[\ln \sigma_0 + T \ln \sigma_e + T \ln \sigma_w + (T + 0.5)(\ln 2\pi + 1)]. \quad (3.37)$$

Therefore, the value of likelihood is mainly decided by sample size, process noises and observation noises. When T is large enough, we can assume

$$\lim_{T \rightarrow \infty} \frac{E(\Omega)}{T} = -\ln \sigma_e - \ln \sigma_w - \ln 2\pi. \quad (3.38)$$

$\lim_{T \rightarrow \infty} \frac{E(\Omega)}{T}$ can be treated as the average joint likelihood for each time instant with input and output sample set of the time instant. The absolute value of the likelihood will be almost proportional to the sample size T .

As we know, the variance of an maximum likelihood estimator, Ψ or $\hat{\Psi}^{\text{ML}}$, is calculated by the inverse of the information matrix, $I(\Psi)$, which is the inverse of the negative of the expected value of the Hessian matrix:

$$H(\Psi) = \frac{\partial^2 \log P(\{\mathbf{x}\}, \{\mathbf{y}\})}{\partial \Psi \partial \Psi'}. \quad (3.39)$$

The variance-covariance matrix of $\hat{\Psi}^{\text{ML}}$ is

$$\text{Var}(\Psi) = [I(\Psi)]^{-1} = -(E[H(\Psi)])^{-1} = \left(-E \left[\frac{\partial^2 \log P(\{\mathbf{x}\}, \{\mathbf{y}\})}{\partial \Psi \partial \Psi'} \right] \right)^{-1}. \quad (3.40)$$

As we'll see shortly, the standard errors of the estimator, $\hat{\Psi}^{\text{ML}}$, are just the square roots of the diagonal terms in the variance-covariance matrix.

According to Cramer-Rao theorem, the MLE is an efficient estimate. When the sample size is large enough, the asymptotic variances of the estimates can be considered as the metric of the accuracy of the estimation. The vector of the asymptotic variances of the estimates is

$$\begin{bmatrix} \frac{(T+1)(\alpha^2-1)^2 U \sigma_e^2}{(T+1)(1-\alpha^2)^2 (P_0 U - U \sigma_0^2 - M_0^2) + (1+T-\alpha^2+T\alpha^2) U \sigma_e^2} \\ \frac{[(1+T)(1-\alpha^2)^2 (P_0 - \sigma_0^2) + (1+T-\alpha^2+T\alpha^2) \sigma_e^2] \sigma_e^2}{(T+1)(1-\alpha^2)^2 (P_0 U - U \sigma_0^2 - M_0^2) + (1+T-\alpha^2+T\alpha^2) U \sigma_e^2} \\ \frac{[(1-\alpha^2)^2 (P_0 U - M_0^2 - U \sigma_0^2) + (1+\alpha^2) U \sigma_e^2] \sigma_e^2}{(T+1)(1-\alpha^2)^2 (P_0 U - U \sigma_0^2 - M_0^2) + (1+T-\alpha^2+T\alpha^2) U \sigma_e^2} \\ \frac{\sigma_w^2}{2T} \\ \sigma_0^2 \end{bmatrix}.$$

When T is large than 1, as shown below, the estimation accuracy for μ_0 will not change, while the accuracy for σ_e and σ_w will be improving. For α and γ , the changing of the accuracy depends on stationary of the time series.

$$\begin{bmatrix} \frac{(\alpha^2 - 1)^2 U v_e}{(1 - \alpha^2)^2 (P_0 U - U v_0 - M_0^2) + (1 + \alpha^2) U v_e} \\ \frac{[(1 - \alpha^2)^2 (P_0 - v_0) + (1 + \alpha^2) v_e] v_e}{(1 - \alpha^2)^2 (P_0 U - U v_0 - M_0^2) + (1 + \alpha^2) U v_e} \\ \frac{[(1 - \alpha^2)^2 (P_0 U - M_0^2 - U v_0) + (1 + \alpha^2) U v_e] v_e}{T(1 - \alpha^2)^2 (P_0 U - U v_0 - M_0^2) + (1 + \alpha^2) U v_e} \\ \frac{\sigma_w^2}{2T} \\ \sigma_0^2 \end{bmatrix}.$$

Per Cramer-Rao Theorem, given certain regularity conditions concerning the distribution, the variance of any unbiased estimator of a parameter, Φ , must satisfy:

$$Var(\Psi) \geq -(E[H(\Psi)])^{-1}.$$

This means that any unbiased estimator that achieves this lower bound is efficient and no better unbiased estimator is possible. Since, $Var(\Psi) = -(E[H(\Psi)])^{-1}$, i. e., no consistent estimator has lower asymptotic mean squared error (MSE) than the of the MLE, MLE is an efficient estimator.

If the sample size is small, we introduce boot-strapping procedure where the estimates are obtained from likelihood constructed from re-sampled standardized innovation, \hat{y}_t , in Eq. 3.21c. Moreover, the mean squared errors of the state variables which is estimated from Eqs. 3.21a-g using estimated parameters are also estimated.

3.4 Data Generation and Simulation Results

The output data is generated through presetting the input signal and the values of the parameters. We implement the initial guess, the EM algorithm and so on, and finally obtain the estimates of the parameters. This makes it easy to evaluate our work by comparing the actual values with the estimated values, or by checking the standard deviation of the estimates.

3.4.1 Data generation

We generate data from the state space model described as Eq. 3.1. We assume $\alpha = 0.8$, $\gamma = 1.5$, and $\mu_0 = 0$. Moreover, the process noises, e_t , and observation noise, w_t , are generated independently where $e_t \sim N(0, 1.1^2)$ and $w_t \sim N(0, 0.9^2)$. We assume that the input, u_t , is a slow changing periodical square wave signal whose period is 10 time units. The standard deviation of initial state, σ_0 , is not needed during the data generation but can be calculated according to Eq. 3.2. The expected log-likelihood can be calculated using Eq. 3.27. Both are treated as “actual” values to be compared with guessed values and estimated values.

We performed our simulation using two different sample sizes: the large size of 1000 and the small size of 50. When the sample size is small, we apply the bootstrapping method to estimate the accuracy of the estimates.

3.4.2 Results

We provide the results of the simulation with the small sample size of 50 in Table 3.1, and the results of the simulation with the large sample size of 1000 in Table 3.2.

In Table 3.1 and Table 3.2, we display the actual values, the guessed values, the estimated values and the standard deviations of the estimated values for transition coefficient, α , input coefficient, γ , standard deviation of process errors, σ_e , standard

Table 3.1: The parameters obtained after three iterations (Sample Size: 50)

Parameters	Actual	Guessed	Estimated	Std. Dev.
α	0.8	0.879	0.801	0.023
γ	1.5	1.075	1.438	0.104
σ_e	1.1	1.400	0.708	0.158
σ_w	0.9	0.922	0.862	0.193
μ_0	0	0.992	2.143	1.231
σ_0	1.83	3.811	1.183	0.252

Table 3.2: The parameters estimated with large sample size (Sample Size: 1000)

Parameters	Actual	Guessed	Estimated	Std. Dev.
α	0.8	0.879	0.800	0.0001
γ	1.5	1.276	1.424	0.0012
σ_e	1.1	2.024	1.033	0.0011
σ_w	0.9	0.533	0.971	0.0005
μ_0	0	2.459	1.667	2.9614
σ_0	1.83	4.250	1.721	NA

deviation of observation errors, σ_w , mean of initial state, μ_0 , and standard deviation of initial state, σ_0 . The standard deviations in Table 3.1 are from the bootstrapped distribution while the standard deviations in Table 3.2 are from the asymptotic variances. In general, the guessed values are near the ‘actual’ values while the estimated values are much closer to the ‘actual’ values than the guessed ones, especially for the parameters of interest: α and

γ . It is worth noting that the deviations of the estimated values are larger than the asymptotic ones due to the imperfect generated data.

3.5 Conclusions

The purpose of this study in this chapter is to try new methods in the application of the EM algorithm in the MLE of a constrained dynamic linear system with exogenous input. There are three main contributions in this chapter. Firstly, we realize that the stand σ_0 and σ_e has the relationship expressed by Eq. 3.4 thus we don't have to estimate them during the implementation of the EM algorithm. Accordingly, the likelihood function in Eq. 3.27 is not like those provided by previous researchers who overlooked the relationship. Secondly, in initial guess of the value of control coefficient and the mean of initial state, we introduce weighted least square for the guess of control coefficient, γ , and the mean of the initial state, μ_0 . Thirdly, since the estimation of a state space model with exogenous time-varying inputs are not found in time series analysis and econometrics, the research in this chapter is an important attempt. The success in the estimation means that we are ready to estimate state space models which are more complicated than the model proposed in this chapter.

It is obvious that more techniques must be found for both the initial guessing and the estimating based on the initial guess, especially during the implementation of the M step of the EM algorithm. The approaches we proposed can be a new start point for the future research on the estimation of dynamic systems with higher dimensions of exogenous inputs, hidden states and observations. Some of these ideas will be used in Chapter IV.

3.6 Chapter Summary

The EM algorithm is popularly used in maximum likelihood estimation of parameters for state space models. However, extant approaches for the realization of the EM algorithm are still not able to fulfill the task of identification dynamic systems which have exogenous inputs. In this chapter, for the state space model representing a system with an exogenous input series, we propose new approaches for the initial guessing of parameter values and the MLE of these values. Using WLS for the initial guessing and the partial differentiation of the joint log-likelihood function for the EM algorithm, we not only estimated the parameters but also compared the estimated values with the “actual” values, which are set to generate simulation data. Moreover, asymptotic variances of the estimated parameters are calculated when the sample size is large, while distributions of the estimated parameters are obtained through bootstrapping when the sample size is small. The results show that the estimated values are close to the “actual” values. This indicates the approaches used are promising in MLE of a state space model with a exogenous input series.

Chapter IV

The Estimation of a State Space

Model with Partly Known

Parameters and Nonlinear

Constraints

In this chapter, we continue to apply state space models to model the dynamics of brand equity. Compared with the ARX models used in Chapter II when sample size is small and the brand values are unknown, state space models are proposed when the sample size is large and brand values are unknown. Compared with the state space model used in Chapter III, the state space model proposed in this chapter has two time-varying inputs and nonlinear constraints, thus is useful to explain a system about brand equity and to study the

effects of different components in this system. However, we must recognize that it is much more challenging to estimate such models than to estimate the model used in Chapter III. As a result, we will concentrate on the design and development of new approaches and the application of existing methods used in areas other than marketing one. As mentioned in Chapter I, the application of the proposed model to empirical study will be future work to be completed when enough data about a firm's brand performance, expenditures on advertising and R&D are available.

4.1 A State Space Model for Brand Equity Dynamics

State space models are widely used in various applications. In economics, a financial model assumes that the trend, seasonal and cycles are hidden stochastic processes and the observations are produced through their integration [95]. In environmental research, a linear Gaussian state space model is applied to sea surface temperature [96]. In marketing research, a state space model is employed to investigate sales response in a certain good category [97] to show how marketing activities can be important in building new brands and in managing existing brands [98], or to investigate the dynamic effects of advertising and word of mouth on the demand for a new product [99].

Denote \mathbf{u}_t as the vector of inputs, \mathbf{x}_t as the vector of hidden states, and \mathbf{y}_t as the vector of outputs at time t . A state space model is described with the process equation (4.1.1) and the observation equation (4.1.2):

$$\mathbf{x}_t = \Phi_t \mathbf{x}_{t-1} + \Gamma_t \mathbf{u}_t + \mathbf{e}_t, \quad (4.1.1)$$

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{w}_t, \quad (4.1.2)$$

where \mathbf{e}_t is the vector of process noises assumed to be normally distributed with mean, $\mathbf{0}$, and covariance, \mathbf{Q}_t , i.e., $\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{Q}_t)$. \mathbf{w}_t is the vector of observation noises assumed to be normally distributed with mean, $\mathbf{0}$, and covariance, \mathbf{R}_t , i.e., $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{R}_t)$. Moreover, Φ_t is the state transition matrix; Γ_t is the input matrix; \mathbf{H}_t is the design matrix or observation matrix. Furthermore, the initial state, $\mathbf{x}_0 \sim N(\boldsymbol{\mu}_0, \mathbf{V}_0)$, and the noise vectors at each step, $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_t\}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_t\}$, are all assumed to be mutually independent.

The hidden states can be estimated through Kalman filtering [100] and Kalman smoothing in Sect. 4.2.

4.1.1 A state space model for brand equity dynamics

An application of state space model is on the structural analysis of brand performance [101] [102] through brand label and brand operation. Brand label is an important firm asset associated with customer brand recognition and perception. Brand label and its associate customer knowledge have value, which we call brand label value (BLV). This value contributes to brand performance value, but is more persistent than brand performance. Brands generate performance through brand operation, which involves the key role of an organization. A brand will generate more earnings if the organization allows better market space, distribution, brand synergy and other infrastructure supports. We call the value generated through brand leverage brand operation value (BOV) and examine BOV through firm investments and its capabilities to leverage these investments.

In the dynamic of brand performance, we define the observation matrix as known, i.e., $H = [1 \quad 1]^T$. The components in system matrix and the input matrix are partly known,

i.e., $\Phi = \begin{bmatrix} \alpha & \beta \\ 0 & 0 \end{bmatrix}$ and $\Gamma = \begin{bmatrix} 0 & 0 \\ \gamma_1 & \gamma_2 \end{bmatrix}$. The observation covariance matrix is actually degenerated as a scalar, $\mathbf{R} = r$. In addition, regarding the initial state, \mathbf{x}_0 , we define its mean as, $\boldsymbol{\mu}_0 = \begin{bmatrix} \mu_{10} \\ \mu_{20} \end{bmatrix}$, and its covariance as, $\mathbf{V}_0 = \begin{bmatrix} v_{10} & v_{x0} \\ v_{x0} & v_{20} \end{bmatrix}$. Moreover, we further define, at time t ,

- $\mathbf{y}_t = BP_t$, where BP_t is the brand performance;
- $\mathbf{x}_t = \begin{bmatrix} BLV_t \\ BOV_t \end{bmatrix}$, where BLV_t and BOV_t are brand label value and brand operation value respectively;
- $\mathbf{u}_t = \begin{bmatrix} Adv_t \\ RD_t \end{bmatrix}$, where Adv_t and RD_t are advertising expenses and R&D expenses respectively;
- $\mathbf{e}_t = \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$ where $e_{1,t}$ and $e_{2,t}$ are the process noises with the variance-covariance matrix \mathbf{Q} , $\mathbf{Q} = \begin{bmatrix} q_1 & q_x \\ q_x & q_2 \end{bmatrix}$;
- $\mathbf{w}_t = w_t$ where w_t is the observation noise with variance r .

From the above brand dynamic formulation, we have the scalar form of the BE dynamic model:

$$BLV_t = \alpha BLV_{t-1} + \beta BOV_{t-1} + e_{1t}, \quad (4.1.3)$$

$$BOV_t = \gamma_1 Adv_t + \gamma_2 RD_t + e_{2t}, \quad (4.1.4)$$

$$BP_t = BLV_t + BOV_t + w_t, \quad (4.1.5)$$

i.e., BOV is created through firm's investments, and its capabilities to leverage these investments. Brand operation will contribute to BLV through brand performance.

In order to be distinguished from the general state space model as described in Eqs. 4.1.1 and 4.1.2, we call the model with its defined matrices, Φ, H, Γ, Q and R, μ_0 and V_0 , as a parsimonious state space model.

In this research, we intend to estimate the unknown elements of the matrices, Φ, H, Γ, Q and R, μ_0 and V_0 . Since the values of these matrices are partly known, our task is to estimate the vector, ψ , or ψ_{13} in this section, which is from a set of 13 unknown parameters, or elements of these matrices:

$$\psi = \psi_{13} = [\alpha, \beta, \gamma_1, \gamma_2, q_1, q_x, q_2, r, \mu_{10}, \mu_{20}, v_{10}, v_{x0}, v_{20}]. \quad (4.1.6)$$

These parameters are constrained by certain linear or nonlinear equations, such as Eq. 4. 5.15 in Sect. 4.5. Generally, we denote the i th constraint as:

$$h_i(\psi) = 0. \quad (4.1.7)$$

In practice, α, β, γ_1 and γ_2 are usually the parameters of interest among all 13 parameters. This implies that the accuracy of the estimates of these four parameters is often emphasized. Therefore, we expect accurate estimation on these four parameters. When necessary, it is acceptable to sacrifice the accuracy of the estimated values of other parameters to guarantee the accuracy of the estimates of the parameters of the interest.

4.1.2 Characteristics of the parsimonious model

The parsimonious state space model has four characteristics: (1) linear Gaussian model, (2) time-variant exogenous input series, (3) partly known parameters, (4) nonlinear constraints to components of matrices.

Firstly, without considering nonlinear constraints, the parsimonious model is linear Gaussian. The estimation of linear state space model has been studied several researchers [103] [89] [104] [105] and the estimation of nonlinear state space model has been studied

by Schon et al. [106] and Kokkala et al. [107]. However, the models studied in these papers do not have the characteristics of this parsimonious model. For example, all those models are fully parameterized.

Secondly, the model includes time-variant exogenous inputs while some elements of the input matrix must be estimated. In previous works [95] [96] [103] [89] [107], linear state space models without inputs are estimated. Although Ghahramani and Hinton [89] mention the possibility of inputs in the model and state that the extension is straightforward, only Gibson and Ninness [105] propose an estimation method to fully parametrized state space models of linear dynamic systems with exogenous input series.

Thirdly, the model is not fully parameterized. Some or all elements in the system matrix, Φ_t , the input matrix, Γ_t , and the design matrix, H_t , of the state space model are known. In contrast, most state space models, such as those used by many authors [89] [103] - [107], to be estimated are fully parametrized models where all elements of all matrices are unconstrained. Fully parametrized state space models provide a very general, compact and simple framework to represent finite-dimensional multivariable systems [105]. However, those methods are not applicable to identify a dynamic system [106] represented by state space models, as in [95] - [99], where some of the parameters are known. The estimation methods for such models must be developed.

Fourthly, there are nonlinear constraints to certain parameters, especially to the components of matrices for the parsimonious model. Linear constraints to the transition matrix, Φ , are investigated [103] [104]. However, as we know, the state space models with nonlinear constraints have not been investigated in the extant literature.

In this chapter, we intend to identify a state space model with the above four characteristics. To limit the complexity of our research, we assume the matrices for the state space model are time-invariant. In other words, our purpose is to find the unknown elements of the matrices, Φ , Γ , H , Q and R , μ_0 and V_0 .

In extant literature, as we know, no methods have been provided to estimate the parameters of a dynamic system with all four characteristics. However, some proposed approaches in extant literature are available to a dynamic system with some of these four characteristics. Roughly, these methods can be sorted into two categories. The first category is based on maximum likelihood estimation (MLE) [108] while the second category is based on Markov chain Monte Carlo (MCMC) [109] [110]. They have their own advantages and disadvantages. In the following two sections, we will introduce two popular methods to estimate the parameter of a state space model: MLE and MCMC. We intend to develop our proposed method after comparing these two.

This chapter is organized as follows. In Sect. 4.2 and Sect. 4.3, we respectively introduce the two categories of methods: MLE and MCMC (through literature review). In Sect. 4.4, we compare these two categories of methods and propose our approach for guessing and estimating the parameters of a dynamic system. Our proposal emphasizes two crucial steps: the guessing of starting value and the integration of methods for nonlinear optimization. The procedure to guess the starting values for the EM algorithm is detailed in Sect. 4.5. The integration of methods for nonlinear optimization in the M step of the EM algorithm is presented in Sect. 4.6. The simulation and its results are presented in Sect. 4.7. Finally, the discussion, conclusion and future work are given in Sect. 4.8.

4.2 Existing Methods: MLE

In statistics, MLE is a method of estimating the parameters of a statistical model given observations, by finding the parameter values that maximize the likelihood of making the observations given the parameters. To introduce the application of MLE in the estimation of state space models, we need to introduce the Kalman filter and the Kalman smoother first.

4.2.1 Kalman filter and Kalman smoother

Assuming we already know the parameter values from the $(k-1)$ th iteration, or the starting values of these parameters, as well as the observations, y_t , and inputs, \mathbf{u}_t , we can estimate the hidden states and their variances at time t based on the observations from steps 1 to t using the Kalman filtering algorithm as shown though Eqs. 4.2.1 - 4.2.7, where the parameter values are those from $k-1$ th iteration, or the starting value when $k = 1$.

$$\mathbf{x}_{t|t-1} = \Phi \mathbf{x}_{t-1|t-1} + \Gamma \mathbf{u}_t, \quad (4.2.1)$$

$$\mathbf{V}_{t|t-1} = \Phi \mathbf{V}_{t-1|t-1} \Phi' + \mathbf{Q}, \quad (4.2.2)$$

$$\tilde{y}_t = y_t - \mathbf{H} \mathbf{x}_{t|t-1}, \quad (4.2.3)$$

$$\Sigma_t = \mathbf{H} \mathbf{V}_{t|t-1} \mathbf{H}' + \mathbf{R}, \quad (4.2.4)$$

$$\mathbf{K}_t = \mathbf{V}_{t|t-1} \mathbf{H}' \Sigma_t^{-1}, \quad (4.2.5)$$

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}_t \tilde{y}_t, \quad (4.2.6)$$

$$\mathbf{V}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{H}) \mathbf{V}_{t|t-1}, \quad (4.2.7)$$

where $\mathbf{x}_{0|0} = \boldsymbol{\mu}_0 = \begin{bmatrix} \mu_{1,0} \\ \mu_{2,0} \end{bmatrix}$ and $\mathbf{V}_{0|0} = \mathbf{V}_0$. \tilde{y}_t is called as “innovation”.

We extend the algorithm proposed by Shumway and Soffer [103] to a dynamic model with input $\mathbf{u}(t)$, to compute the expectation $E[\mathbf{x}_t|\{y, u\}_1^T] \equiv \mathbf{x}_t^T$ and the correlation matrices $\mathbf{P}_t \equiv \mathbf{V}_t^T + \mathbf{x}_t^T(\mathbf{x}_t^T)'$ based on past and future observation data and input. A set of backward recursions is performed using Eqs. 4.2.8 - 4.2.11:

$$\mathbf{J}_{t-1} = \mathbf{V}_{t-1|t-1} \boldsymbol{\Phi}' [\mathbf{V}_{t|t-1}]^{-1}, \quad (4.2.8)$$

$$\mathbf{x}_{t-1}^T = \mathbf{x}_{t-1|t-1} + \mathbf{J}_{t-1}(\mathbf{x}_t^T - \mathbf{x}_{t-1|t}), \quad (4.2.9)$$

$$\mathbf{V}_{t-1}^T = \mathbf{V}_{t-1|t-1} + \mathbf{J}_{t-1}(\mathbf{V}_t^T - \mathbf{V}_{t|t-1})\mathbf{J}_{t-1}', \quad (4.2.10)$$

where $\mathbf{x}_T^T = \mathbf{x}_{T|T}$ and $\mathbf{V}_T^T = \mathbf{V}_{T|T}$.

We also have $\mathbf{P}_{t,t-1} \equiv \mathbf{V}_{t,t-1}^T + \mathbf{x}_{t-1}^T(\mathbf{x}_{t-1}^T)'$, which can be obtained through the backward recursions

$$\mathbf{V}_{t-1|t-2}^T = \mathbf{V}_{t-1|t-1}\mathbf{J}_{t-2}' + \mathbf{J}_{t-1}(\mathbf{V}_{t|t-1}^T - \boldsymbol{\Phi}\mathbf{V}_{t-1|t-1})\mathbf{J}_{t-2}', \quad (4.2.11)$$

where $\mathbf{V}_{t-1|t-2}^T$ is initialized using $\mathbf{V}_{T|T-1}^T = (\mathbf{I} - \mathbf{K}_T\mathbf{H})\boldsymbol{\Phi}\mathbf{V}_{T-1|T-1}$.

4.2.2 Mathematical optimization problem

The estimation of the parameters of the state space model, as in Eqs. 4.1.1 and 4.1.2, is quite complicated. We use $\boldsymbol{\psi}_{13}$ to represent the vector of parameters containing the elements of the initial mean and covariance $\mu_{10}, \mu_{20}, \nu_{10}, \nu_{20}$, and ν_{x0} , the transition coefficients, α and β , the input components Y_1 and Y_2 , and the elements, q_1, q_2, ν_{10}, q_x , of the state and observation covariance matrices \mathbf{Q} and \mathbf{R} . We use maximum likelihood under the assumption that the initial state is normal, $\mathbf{x}_0 \sim \mathbf{N}(\boldsymbol{\mu}_0, \mathbf{V}_0)$, and the errors w_1, \dots, w_n and ν_1, \dots, ν_n are jointly normal and uncorrelated vector variables. We continue to assume, for simplicity, that $\{w_t\}$ and $\{\nu_t\}$ are uncorrelated.

The value of the log likelihood function can be calculated at each stage using the “innovations” form, i.e., Eq. 4.2.3.

The innovations form of the likelihood function proceeds by noting the innovations are independent Gaussian random vectors with zero means and, covariance matrices as shown in Eq. 4.2.4.

Hence, ignoring a constant and multiplying a number -2, we may write the doubled negative log likelihood, $D_y(\boldsymbol{\psi})$, as

$$D_y(\boldsymbol{\psi}) = \log|\mathbf{V}_t| + \sum_{t=1}^T \ln(\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_{t|t-1})^T \mathbf{V}_t^{-1} (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_{t|t-1}), \quad (4.2.12)$$

where we have emphasized the dependence of the innovations on the parameters $\boldsymbol{\psi}$. Accordingly, the mathematical optimization problem is

$$\max_{\boldsymbol{\psi}} -D_y(\boldsymbol{\psi}), \quad (4.2.13)$$

or

$$\min_{\boldsymbol{\psi}} D_y(\boldsymbol{\psi}). \quad (4.2.14)$$

4.2.3 Algorithm for MLE: Newton-Raphson

$D_y(\boldsymbol{\psi})$ is a highly nonlinear and complicated function of the unknown parameters. The usual procedure is to fix \mathbf{x}_0 and then develop a set of recursions for $D_y(\boldsymbol{\psi})$ and its first two derivatives. Then, a Newton–Raphson algorithm can be applied repeatedly to update the parameter values until $D_y(\boldsymbol{\psi})$ is minimized. The steps involved in performing a Newton–Raphson estimation procedure are as follows.

- (i) Select starting values, $\boldsymbol{\psi}^{(0)}$, for the parameters.

- (ii) Run the Kalman filter, Eq. 4.2.1 - 4.2.7, using the starting parameter values, $\boldsymbol{\psi}^{(0)}$, to obtain a set of innovations and error covariances, $\{\tilde{y}_t^{(0)} ; t = 1, \dots, T\}$ and $\{\mathbf{V}_t^{(0)} ; t = 1, \dots, T\}$.
- (iii) Run one iteration of a Newton–Raphson procedure with $D_y(\boldsymbol{\psi})$ as the criterion function, to obtain a new set of estimates, say $\boldsymbol{\psi}^{(1)}$.
- (iv) At iteration i ($i = 1, 2, \dots$), repeat step (ii) using $\boldsymbol{\psi}^{(i)}$ in place of $\boldsymbol{\psi}^{(i-1)}$ to obtain a new set of innovation values $\{\tilde{y}_t^{(0)} ; t = 1, \dots, T\}$ and $\{\mathbf{V}_t^{(i)} ; t = 1, \dots, T\}$. Then repeat step 3 to obtain a new estimate $\boldsymbol{\psi}^{(i+1)}$.
- (v) Stop when the estimates or the likelihood stabilize; for example, stop when the values of $\boldsymbol{\psi}^{(i+1)}$ differ from $\boldsymbol{\psi}^{(i)}$, or when $-\log L(\boldsymbol{\psi}^{(i+1)})$ differs from $-\log L(\boldsymbol{\psi}^{(i)})$, by some predetermined, but small amount.

4.2.4 Algorithm for MLE: EM

Generally, the MLE methods have a few undesired features which can be avoided using the EM (Expectation-Maximization) algorithm [88]. Firstly, the calculation of the inverse of the matrix of second order partials can be fairly large if there is a noticeable number of parameters. As a result, the corrections in the successive iterations is generally quite involved. Secondly, it is not guaranteed that the successive steps in Newton-Raphson algorithm will decrease $D_y(\boldsymbol{\psi})$. Sometimes, if an extremely large step is encountered, $D_y(\boldsymbol{\psi})$ may even increase.

On the other hand, the EM steps always decrease $D_y(\boldsymbol{\psi})$. For an exponent family, convergence to a stationary point is guaranteed. Depending on the shape of $D_y(\boldsymbol{\psi})$, a local or global minimum can be found. In contrast to the highly nonlinear appearance of the

Newton-Raphson expressions, the EM equations usually have a simple heuristically appealing form. When there are no inputs in the model in the form of Eqs. 4.1.1 and 4.1.2. The basic idea is that if we could observe the states at time t , $\mathbf{x}_{t:T} = \{x_0, x_1, \dots, x_T\}$, in addition to the observations, $\mathbf{y}_{t:T} = \{y_0, y_1, \dots, y_T\}$, then we would consider $\{\mathbf{x}_{t:T}, \mathbf{y}_{t:T}\}$ as the complete data, with the joint density

$$f_{\boldsymbol{\psi}}(\mathbf{x}_{t:T}, \mathbf{y}_{t:T}) = f_{\mu_0, V_0}(\mathbf{x}_0) \prod_{t=1}^T f_{\boldsymbol{\psi}, Q}(\mathbf{x}_t | \mathbf{x}_{t-1}) \prod_{t=1}^T f_R(\mathbf{y}_t | \mathbf{x}_t). \quad (4.2.15)$$

Under the Gaussian assumption and ignoring constants, the doubled negative log-likelihood from Eq. 4.2.15 with complete data $\{\mathbf{x}_t, \mathbf{y}_t\}$, can be written as

$$\begin{aligned} \mathbf{D}_{x,y}(\boldsymbol{\psi}) = & \ln|V_0| + (\mathbf{x}_0 - \boldsymbol{\mu}_0)' V_0^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}_0) + T \ln|Q| + \sum_{t=1}^T (\mathbf{x}_t - \\ & \boldsymbol{\Phi} \mathbf{x}_{t-1})' Q^{-1} (\mathbf{x}_t - \boldsymbol{\Phi} \mathbf{x}_{t-1}) + T \ln|R| + \sum_{t=1}^T (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_{t-1})' R^{-1} (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_{t-1}). \end{aligned} \quad (4.2.16)$$

Since we actually do not have the complete data, we apply the EM algorithm, an iterative method, to find the MLEs of $\boldsymbol{\psi}$ based on the incomplete data, \mathbf{Y}_T , by successively maximizing the conditional expectation of $\mathbf{D}_{x,y}(\boldsymbol{\psi})$.

To implement the EM algorithm, we denote, at iteration i , ($i = 1, 2, \dots$),

$$G(\boldsymbol{\psi}^{(i)} | \boldsymbol{\psi}^{(i-1)}) = E\{\mathbf{D}_{x,y}(\boldsymbol{\psi}) | \{\mathbf{y}_t\}, \boldsymbol{\psi}^{(i-1)}\}. \quad (4.2.17)$$

Calculation of Eq. 4.2.17 is the expectation step. Given the current value of the parameters, $\boldsymbol{\psi}^{(j-1)}$, we can apply the Kalman smoothing (Eqs. 4.2.8 - 4.2.11) and then obtain the desired conditional expectations. We have

$$\begin{aligned} G(\boldsymbol{\psi}^{(i)} | \boldsymbol{\psi}^{(i-1)}) = & \ln|V_0| + \text{tr}\{V_0^{-1} [\mathbf{P}_0^T + (\mathbf{x}_0^T - \boldsymbol{\mu}_0)(\mathbf{x}_0^T - \boldsymbol{\mu}_0)']\} + T \ln|Q| + \\ & \text{tr}\{Q^{-1} [\mathbf{S}_{11} - \mathbf{S}_{10} \boldsymbol{\Phi}' - \boldsymbol{\Phi} \mathbf{S}_{10}' + \boldsymbol{\Phi} \mathbf{S}_{00}' \boldsymbol{\Phi}']\} + T \ln|R| + \text{tr}\{R^{-1} \sum_{t=1}^T [(\mathbf{y}_t - \\ & \mathbf{A}_t \mathbf{x}_t^T)(\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t^T)' + \mathbf{A}_t \mathbf{P}_t^T \mathbf{A}_t']\}, \end{aligned} \quad (4.2.18)$$

where

$$\mathbf{S}_{11} = \sum_{t=1}^T (\mathbf{x}_t^T \mathbf{x}_t^{T'} + \mathbf{P}_t^T), \quad (4.2.19)$$

$$\mathbf{S}_{10} = \sum_{t=1}^T (\mathbf{x}_t^T \mathbf{x}_{t-1}^{T'} + \mathbf{P}_{t,t-1}^T), \quad (4.2.20)$$

and

$$\mathbf{S}_{00} = \sum_{t=1}^T (\mathbf{x}_{t-1}^T \mathbf{x}_{t-1}^{T'} + \mathbf{P}_{t-1}^T). \quad (4.2.21)$$

Minimizing Eq. 4.2.18 with respect to the parameters, at iteration i , is the maximization step of the EM algorithm, which yields the updated estimates

$$\boldsymbol{\Phi}^{(i)} = \mathbf{S}_{10} \mathbf{S}_{00}^{-1}, \quad (4.2.22)$$

$$\mathbf{Q}^{(i)} = T^{-1} (\mathbf{S}_{11} - \mathbf{S}_{10} \mathbf{S}_{00}^{-1} \mathbf{S}_{10}'), \quad (4.2.23)$$

and

$$\mathbf{R}^{(i)} = T^{-1} \sum_{t=1}^T [(\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t^T)(\mathbf{y}_t - \mathbf{A}_t \mathbf{x}_t^T)' + \mathbf{A}_t \mathbf{P}_t^T \mathbf{A}_t']. \quad (4.2.24)$$

The updates for the initial mean and variance–covariance matrix are

$$\boldsymbol{\mu}_0^{(i)} = \mathbf{x}_0^T, \quad (4.2.25)$$

and

$$\mathbf{V}_0^{(i)} = \mathbf{P}_0^T. \quad (4.2.26)$$

obtained from minimizing Eq. 4.2.18.

The iterative procedure of the EM algorithm is as follows.

- (i) Initialize the procedure by selecting starting values for the parameters $\boldsymbol{\psi}^{(0)} = \{\boldsymbol{\mu}_0, \mathbf{V}_0, \boldsymbol{\Phi}, \mathbf{Q}, \mathbf{R}\}$.

On iteration i , ($i = 1, 2, \dots$):

- (ii) Compute the incomplete-data likelihood, $D_{\mathbf{y}}(\boldsymbol{\psi}^{(i-1)})$; see Eq. 4.2.12.

- (iii) Perform the E Step. Obtain the smoothed values \mathbf{x}_t^T , \mathbf{P}_t^T and $\mathbf{P}_{t,t-1}^T$, for $t = 1, \dots, T$, using the parameters $\boldsymbol{\psi}^{(i-1)}$. Use the smoothed values to calculate \mathbf{S}_{11} , \mathbf{S}_{10} , \mathbf{S}_{00} given in Eqs. 4.2.19 - 4.2.21.
- (iv) Perform the M Step. Update the estimates, $\{\boldsymbol{\mu}_0, \mathbf{V}_0, \boldsymbol{\Phi}, \mathbf{Q}, \mathbf{R}\}$ using Eqs. 4.2.22 - 4.2.26, to obtain $\boldsymbol{\psi}^{(i)}$.
- (v) Repeat steps (ii) – (iv) to convergence.

4.3 Existing Method: Markov Chain Monte Carlo

In statistics, MCMC methods are a class of sampling algorithms for a probability distribution. The sampling algorithm is based on constructing a Markov chain. And the Markov chain has the expected distribution as its equilibrium distribution. After a number of steps, the state of the chain (not the state of a state space model) is then used as a sample of the expected distribution. MCMC based methods are recommended in parameter estimation of various mathematical models, including those state space models with time-variant exogenous input series.

In this section, we assume that a state space model has the structure described by Eqs. 4.1.1 and 4.2.2. This structure is a generalized format of the state space model:

$$\mathbf{x}_t | \{\mathbf{x}_{0:t-1}, \mathbf{y}_{0:t-1}\} \sim p(\mathbf{x}_t | \mathbf{x}_{t-1}, \boldsymbol{\psi}), \quad (4.3.1)$$

$$\mathbf{y}_t | \{\mathbf{x}_{0:t}, \mathbf{y}_{0:t-1}\} \sim p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\psi}). \quad (4.3.2)$$

Throughout this subsection, we denote $\mathbf{x}_{0:t} = (\mathbf{x}_0, \dots, \mathbf{x}_t)$, and write $p(\cdot | \cdot)$ for a generic conditional probability density or mass function. Moreover, we specify $p(\mathbf{x}_0 | \boldsymbol{\psi})$ as an initial distribution. The model is dependent on an unknown multi-dimensional

parameter $\boldsymbol{\psi}$. Conditional on the parameter $\boldsymbol{\psi}$, we assume that the state model is Markov, and the observation \mathbf{y}_t only depends on the state at that time, \mathbf{x}_t , at time instant t .

In this section, our intention is to introduce how to perform Bayesian inference for a state space model given data $\mathbf{y}_{1:T}$. Assuming a prior for the parameters, $p(\boldsymbol{\psi})$, has been specified, the posterior of the parameters $p(\boldsymbol{\psi} | \mathbf{y}_{0:T})$, or in some cases the joint distribution of the state and the parameters, $p(\boldsymbol{\psi}, \mathbf{x}_{0:T} | \mathbf{y}_{0:T})$ will be obtained. From these posterior distributions, we can design an MCMC algorithm using data augmentation [111]. The key to the MCMC algorithm is to design a Markov chain whose state is $(\boldsymbol{\psi}, \mathbf{x}_{0:T})$, and whose stationary distribution is $p(\boldsymbol{\psi}, \mathbf{x}_{0:T} | \mathbf{y}_{0:T})$, considering that the stationary distribution of the MCMC algorithm is available up to proportionality:

$$p(\boldsymbol{\psi}, \mathbf{x}_{0:T} | \mathbf{y}_{0:T}) \propto p(\boldsymbol{\psi})p(\mathbf{x}_0 | \boldsymbol{\psi}) \prod_{t=1}^T p(\mathbf{x}_t | \mathbf{x}_{t-1}, \boldsymbol{\psi}) \prod_{t=0}^T p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\psi}). \quad (4.3.3)$$

Usually moves are designed to update $\boldsymbol{\psi}$ conditional on the current values of $\mathbf{x}_{1:T}$ and then update $\mathbf{x}_{1:T}$ conditional on $\boldsymbol{\psi}$. In Sect. 4.3.1 and Sec. 4.3.2, we introduce the problem of updating the state and then move to update the parameters.

4.3.1 Update the state of a state space model

The simplest approach is called a single-site update, i.e., to update the state $\mathbf{x}_{0:T}$ is to update components of $\mathbf{x}_{0:T}$ one at a time. This move is easy to implement. However, if there is strong temporal dependence in the state process, the moves may lead to slow mixing. In these scenarios, people tend to update blocks of state components, $\mathbf{x}_{t:S}$, or the whole state process $\mathbf{x}_{0:T}$ in a single move.

A. Single-site updates of the state

During single-site updates, an MCMC move will update a single value of the state, \mathbf{x}_t , conditional on all other values of the state process (and on $\boldsymbol{\psi}$). The whole state process is updated through repeatedly applying this move for $t = 0, \dots, T$.

Denoting that $\mathbf{x}_{-t} = (\mathbf{x}_0, \dots, \mathbf{x}_{t-1}, \mathbf{x}_{t+1}, \dots, \mathbf{x}_T)$ as the whole state process excluding \mathbf{x}_t , a single-site update will update \mathbf{x}_t for fixed $\mathbf{x}_{-t}, \boldsymbol{\psi}$. Due to the Markov structure of the model (Eq. 4.3.1 and 4.3.2), the target distribution of such a move, which is the full-conditional distribution $p(\mathbf{x}_t | \mathbf{x}_{-t}, \boldsymbol{\psi}, \mathbf{y}_{0:t})$, can be simplified to $p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t+1}, \boldsymbol{\psi}, \mathbf{y}_t)$ for $t = 1, \dots, T-1$, $p(\mathbf{x}_0 | \mathbf{x}_1, \boldsymbol{\psi}, \mathbf{y}_0)$ for $t = 0$ and $p(\mathbf{x}_T | \mathbf{x}_{T-1}, \boldsymbol{\psi}, \mathbf{y}_T)$ for $t = T$.

B. Block updates for the state

If there is strong dependence in the state-process, the resulting MCMC algorithms can mix slowly. Accordingly, block updates are introduced to update the state at more than one time-point in a single move. Ideally the whole state process would be updated in one move, and in some cases it turns out that this is possible to do from the full-conditional, so that moves are always accepted.

Independence proposal can be possibly used to update jointly a block of state values when it is not possible to update the whole state process from its full conditional. Good independence proposals are possibly available for general state-models. However, particularly for high-dimensional states and models with strong nonlinearities, independence proposals can become challenging.

4.3.2 Updating the parameters

Within the MCMC algorithm, the usual approach to update the parameter, $\boldsymbol{\psi}$, is to update it conditional on the current value of the state path $\mathbf{x}_{0:T}$. This is often simple to complete under two cases. The first case is that the conjugate priors for $\boldsymbol{\psi}$ can be chosen so that the sampling can be started directly from $p(\boldsymbol{\psi} | \mathbf{x}_{0:T}, \mathbf{y}_{0:T})$. The second case is that $\boldsymbol{\psi}$ is of sufficiently low-dimension that efficient independence proposals can be used. In some cases, components or blocks of $\boldsymbol{\psi}$ are updated at a time, rather than updating the whole parameter vector in one attempt.

However, if there is strong correlation between $\boldsymbol{\psi}$ and $\mathbf{x}_{0:T}$, the overall efficiency of the MCMC algorithm can still be poor even if the sampling can be executed from the full-conditional $p(\boldsymbol{\psi} | \mathbf{x}_{0:T}, \mathbf{y}_{0:T})$.

When there is strong dependence between $\boldsymbol{\psi}$ and $\mathbf{x}_{0:T}$, there are two techniques to avoid mixing. The first is to consider a different parameterization. This new parameterization is expected to reduce dependence between the state and the parameter. The second is to use moves that jointly update $\boldsymbol{\psi}$ and $\mathbf{x}_{0:T}$.

For state space models, two possible general parameterizations originally for hierarchical models [112] can be used. The one is centered parameterizations, which is defined by a model where $p(\boldsymbol{\psi} | \mathbf{x}_{0:T}, \mathbf{y}_{0:T}) = p(\boldsymbol{\psi} | \mathbf{x}_{0:T})$. The second one is non-centered parameterizations where a priori $\boldsymbol{\psi}$ and $\mathbf{x}_{0:T}$ are independent.

For those models which can be simulated directly from $p(\mathbf{x}_{0:T} | \boldsymbol{\psi}, \mathbf{y}_{0:T})$, jointly updating of $\boldsymbol{\psi}$ and $\mathbf{x}_{1:T}$ is most easily and commonly implemented through choosing $q(\mathbf{x}'_{0:T} | \boldsymbol{\psi}') = p(\mathbf{x}'_{0:T} | \boldsymbol{\psi}', \mathbf{y}_{0:T})$. The resulting acceptance ratio then simplifies to:

$$\min \left\{ 1, \frac{q(\boldsymbol{\psi} | \boldsymbol{\psi}') p(\boldsymbol{\psi}' | \mathbf{y}_{0:T})}{q(\boldsymbol{\psi}' | \boldsymbol{\psi}) p(\boldsymbol{\psi} | \mathbf{y}_{0:T})} \right\}. \quad (4.3.4)$$

This acceptance ratio does not depend on $\mathbf{x}_{0:T}$ or $\mathbf{x}'_{0:T}$. The marginal chain for $\boldsymbol{\psi}$ is equivalent to a MCMC chain for $p(\boldsymbol{\psi} | \mathbf{y}_{0:T})$ with proposal distribution $q(\boldsymbol{\psi}' | \boldsymbol{\psi})$. Provided that an efficient proposal $q(\boldsymbol{\psi}' | \boldsymbol{\psi})$ can be found, such an MCMC algorithm will always be more efficient than one that updates $\boldsymbol{\psi}$ and $\mathbf{x}_{1:T}$ independently. However, the difficulty with implementing this idea is how to choose $q(\boldsymbol{\psi}' | \boldsymbol{\psi})$.

If there is an efficient independence proposal for $\mathbf{x}_{0:T}$ given $\boldsymbol{\psi}$, a simple extension of this joint updating idea is possible because this proposal could be used as $q(\mathbf{x}'_{0:T} | \boldsymbol{\psi}')$. The efficiency of the resulting algorithm will depend on both the efficiency of $q(\boldsymbol{\psi}' | \boldsymbol{\psi})$ as a proposal for an MCMC that explores $p(\boldsymbol{\psi}' | \mathbf{y}_{0:T})$, and also the closeness of $q(\mathbf{x}'_{0:T} | \boldsymbol{\psi}')$ to $p(\mathbf{x}'_{0:T} | \boldsymbol{\psi}', \mathbf{y}_{0:T})$.

4.3.3 Particle algorithm

When the state space model is nonlinear and/or non-Gaussian, MCMC methods can obviously be used for off-line inference as described in Sect. 4.3.1 and Sect. 4.3.2, but they are impractical for online inference. Even for off-line inference, it can be difficult to build efficient high-dimensional proposal distributions for such algorithms. Consequently, for nonlinear non-Gaussian state space models, particle algorithms are becoming the most successful and popular. The popularity is because (1) they are easy to implement, suitable for parallel implementation [113] and (2) more importantly, have been demonstrated in numerous settings to yield more accurate estimates than the standard alternatives [114] [115].

In most applications, including our research, obtaining the latent state of the state space model of interest also depends on unknown static parameters that need to be estimated from the available data of outputs and, in our research, inputs. In fact, inferring the parameter $\boldsymbol{\psi}$ is often the primary problem of interest. In this context, standard particle methods fail and it is necessary to rely on more sophisticated algorithms.

Nowadays most particle algorithms proposed to estimate parameters in general state space models are inefficient in computation because of the degeneracy problem [114]. Several approaches have been proposed to deal with the degeneracy problem by either adding an artificial dynamic on the parameters [116] or introducing a fixed-lag approximation [117] [118]. These methods can work very well in practice, but it remains unfortunately difficult or impossible to quantify the bias introduced in most realistic applications.

4.3.4 Summary

In Sect. 4.3, we have given an introduction to MCMC methods and particle algorithms for general state space models. Three main issues have been covered. Firstly, if there is strong, or long-range, dependence in the state space model, then an efficient MCMC algorithm will need to update blocks of the state process in a single move. Secondly, strong correlation between the parameters and the state process can lead to slow mixing of MCMC algorithm. To improve mixing, either re-parameterization of the model, or joint updates of the state and the parameters will be needed. Thirdly, it is still difficult or even impossible for recent particle algorithms to assess the bias introduced in most realistic applications.

4.4 Proposed Approach

In this section, we look for a solution appropriate to estimate the parameters of the state space model described in Chapter I. We compare The MLE method and the MCMC method first and make choice between them. Then we provide the mathematical problem to be solved in this research and finally emphasize the importance of the guessing of starting values for the EM algorithm.

4.4.1 MLE vs. MCMC

The dynamic systems under investigation in these papers listed in Table 4.1 have features described below, which are at odd with the parsimonious state space model.

- (1) *No exogenous inputs, not to say time-variant inputs.* Deng and Shen [104] studied a state space model with exogenous inputs. However, those inputs are actually time-invariant parameters to be estimated.
- (2) *Fully parametrized.* All the components in the matrices, Φ , Γ , H , Q and R , are unknown parameters to be estimated. The components of the matrices, μ_0 and V_0 , are usually unknown, but was assumed to be known [104] [107].
- (3) *No extra constraints.* As an option, there can exist linear constraints [103], to the transition matrix, Φ . However, state space models with nonlinear constraints to certain parameters are not studied in extant literature, as we know.

On the other hand, MCMC based methods are proposed in parameters estimation of the state space models with time-variant exogenous input series. Table 4.2 lists the published papers where MCMC based methods are used. The dynamic systems under investigation in these papers have the following features which are similar with the

parsimonious state space model, except that the parsimonious model has extra constraints: (1) Linear, (2) Time-variant input series, (3) Partly known parameters, (4) No extra constraints.

It seems that, with some modification, MCMC based methods might be able to be applied to the estimation of parsimonious model. However, because there are always some residual effects of the starting position, MCMC sampling typically can only approximate the target distribution. More sophisticated MCMC-based algorithms, such as Gibbs sampling, can produce exact samples, at the cost of additional computation and an unbounded running time [119]. Regarding linear dynamic systems, we can take, dynamic linear model (DLM), as an example. DLMs are derived from the state space models.

MCMC based approaches with variations are used to estimate DLMS. According to Leeflang et al. [120], the estimation needs a high computational cost, and may take several hours or days, depending on the dimensionality of the problem. Regarding nonlinear dynamic systems, Kokkala et al. [107] concludes that particle filter, or sequential Monte Carlo, and sigma-point filtering, are of higher computational complexity and theoretical exactness than the extended Kalman filter based direct likelihood approximation.

On the other hand, after comparing inference between using the EM algorithm and Gibbs sampling in hidden Markov models (HMM) of different degrees of complexity regarding model structure, Ryden [121] concludes that, if only a point estimate is needed and if for comparing between the models their maximal likelihood are used only, the EM algorithm is generally the simplest and quickest method.

Table 4.1: State space models estimated using MLE and EM in extant literature

Papers	Models					Initial guessing
Authors (Year)	Linearity	Existence of Exogenous Inputs?	Fully Parametrized?	Existence of Extra Constraints	Initial States and their variance as parameters?	
Shumway and Stoffer (1982) [103]	Linear	No	Yes	Optional	Unknown parameters	Examine different sets
Ghahraman and Hilton (1996) [89]	Linear	No	Yes	No	Unknown parameters	Not mentioned
Deng and Shen (1997) [104]	Linear	Constants	Yes	No	Known constants	Just set the values
Gibson and Ninness (2005) [105]	Linear	Yes	Yes	No	Unknown parameters	Subspace-based
Schon et al. [106]	Nonlinear	Yes	Yes	No	Unknown parameters	Choose randomly
Kokkala et al. (2014)[107]	Nonlinear	No	Yes	No	Known parameters	Choose arbitrarily

Table 4.2: State space models estimated using methods based on MCMC in extant literature

Papers	Models					Estimation Methods
Authors (Year)	Linearity	Existence of Exogenous Inputs?	Fully Parametrized?	Existence of Extra Constraints?	Initial States and their variance	
Van Heerde et al. (2004) [97]	Linear	Yes	Partly parametrized	No	Not mentioned	Gibbs sampling
Ataman et al. (2007) [98]	Linear	Yes	Observation Matrix is known	No	Not mentioned	Gibbs sampling and random walk Metropolis-Hastings algorithm
Bruce and Foutz (2007) [99]	Linear	Yes	Partly parametrized	No	Not mentioned	MCMC combining with Kalman filtering and smoothing

In the research on BE dynamics, only the point estimate of the parameters of a state space model, a special type of HMM, is needed. Therefore, to reduce the computational complexity and time-consumption, we aim to apply the MLE method and the EM algorithm, instead of MCMC based methods, in the identification of a dynamic system with time-varying exogenous inputs and partly known matrices. The Kalman filtering and the Kalman smoothing will be applied in the M step of the EM algorithm.

4.4.2 Mathematical problem and its characteristics

Since the model for BE dynamics is linear Gaussian, we use the state space model to illustrate our proposal.

As mentioned in Sect. 4.1, our proposed system has constraints, which reshapes the mathematical optimization problem as below.

Considering s constraints, the mathematical optimization problem formulated is:

$$\min_{\boldsymbol{\psi}} D_{\mathbf{y}}(\boldsymbol{\psi}) \quad (4.4.1)$$

subjected to

$$\begin{cases} h_1(\boldsymbol{\psi}) = 0 \\ h_2(\boldsymbol{\psi}) = 0 \\ \dots \\ h_s(\boldsymbol{\psi}) = 0 \end{cases}. \quad (4.4.2)$$

Applying the method of Lagrange multipliers, the mathematical optimization problem becomes

$$\min_{\boldsymbol{\psi}, \boldsymbol{\lambda}} D_{\mathbf{y}}(\boldsymbol{\psi}) + \sum_{i=1}^s \lambda_i h_i(\boldsymbol{\psi}), \quad (4.4.3)$$

where $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_i, \dots, \lambda_s\}$, are s Lagrange multipliers, while $\boldsymbol{\psi}$ is the parameters to be estimated for the state space model we are investigating.

4.4.3 The EM algorithm and starting values

Since some of the elements of the matrices, Φ , Γ , H , Q and R , μ_0 and V_0 , are assumed to be known, when MLE is applied, the objective function formed from log likelihood and the derivatives of the objective functions, for the M step of the EM algorithm, must be expressed with unknown parameters only, i.e., with some of elements of these matrices. Consequently, the objective function in our research is nonlinear, neither convex nor concave. By contrast, the constructed objective functions [89] [103] - [107] are concave because the mathematical derivations are performed with respect to the whole matrices, Φ , Γ , H , Q and R , μ_0 and V_0 , not elements of them. Moreover, the constraints to the mathematical optimization problem are nonlinear. Consequently, in the frame of the MLE method and the EM algorithm, the optimization methods proposed by us in this chapter must adopt methods other than those proposed [89] [103] - [107].

Nevertheless, the general EM algorithm has problems including slow convergence, the need for an appropriate stopping rule that can sense if the algorithm has reached the maximum, and the choice of starting values in order to reach the global maximum in a smaller number of iterations. Many algorithms aiming to speed up the convergence of EM while preserving its simplicity have been proposed [122] - [126]. The problem of choosing an appropriate stopping rule has also obtained much attention in the literature. Several criteria have been recommended and the effect of stopping early has been scrutinized [127].

The third significant problem of the EM algorithm is that sub-optimal MLEs, usually local optima, can be produced because its solution likely relies on its starting values. Slow convergence and dependence on starting values, can be correlated in practice. It is possible that starting from some values leads to a slower convergence rate for the EM

algorithm and that the EM algorithm is terminated before reaching a sensible value of the objective function.

Hence, it is critical to choose the starting values of the parameters properly because the starting value can seriously affect the speed of convergence of the EM algorithm and its capability to locate the global maximum [128]. In the algorithm-based literature, methods such as grid search and moment method have been recommended. These methods are feasible when the size of parameter space is small. However, when the size of the parameter space is large, the speed of convergence is slow [128].

To avoid the third problem, it is critical to find an appropriate method to choose the starting values of the parameters. Unfortunately, in extant literature on parameter estimation, probably the most employed way of initiating the EM algorithm consists of initializing it from random values. Guessing of the starting values of the parameters are usually ignored, as shown in Table 4.1. Most starting values of the parameters, even all of them [104] [106], are randomly chosen. The only exception is that the initial transition parameter value of the state space model is guessed by examining the observed output series [103]. In this chapter, we develop a procedure to guess the starting values of the parameters to be estimated. With low computational cost, the advantage of the guessing procedure develop is its ability to provide relatively accurate starting values for the EM algorithm than random initialization.

Simple random initialization is often outperformed by other strategies, such as grid search and methods of moments (MOM). Grid search is simply an exhaustive searching through a manually specified subset of the hyperparameter space of a learning algorithm. Grid search is employed to set the initial values to estimate a VSTAR (vector

smooth transition autoregressive) model [129]. Overall, the grid search needs another algorithm to obtain a better, hopefully global, optimum. Moreover, grid search suffers from the curse of dimensionality, and is often embarrassingly parallel because typically the hyperparameter settings it evaluates are independent of each other [130]. MOM [131] is introduced to determine starting values for Gaussian mixture model [132]. MOM is also used [96] in choosing starting values of three parameters of a model based on Ornstein-Uhlenbeck process [133]. These methods are feasible when the size of parameter space is small: less than four unknown parameters in extant literature. However, when the size of the parameter space is large, there will be more equations formed from moments. Accordingly, the speed of convergence will be slow.

For mixture models, Biernacki. et al. [134] and Karlis and Xekalaki [128] investigated or recommended methods to choose starting values for the EM algorithm. In contrast, for state space models, there is no such published survey or review on how to determine the starting values of the parameters of these models as mentioned earlier.

Therefore, in this chapter, our emphases are on (1) guessing the starting values of the parameters to be estimated and (2) integrating mathematical optimization methods to estimate the values of the parameters during the application of the MLE method and the EM algorithm. The procedure to guess the starting values and the way to integrate the optimization methods are elaborated in Sect. 4.5 and Sect. 4.6 respectively.

4.5 Guessing of the Starting Values of the Parameters

In this section, we introduce our procedure to guess the starting values of the parameters to be estimated (ψ_{13}) from the input series, $\{u_{1,t}\} = \{u_{1,1}, u_{1,2}, \dots, u_{1,T}\}$ and

$\{u_{2,t}\} = \{u_{21}, u_{22}, \dots, u_{2T}\}$, and the output series, $\{y_t\} = \{y_1, y_2, \dots, y_T\}$, where T is the length of these data series. In extant literature, only statistical indices of the output series are used to guess the initial value of a certain parameter [103]. In this section, we use not only the statistical indices of input series and output series respectively, but also the cross covariance between the input series and the output series. Moreover, we apply linear regression between the output series and the lagged input series, which is not seen in extant literature.

The development of the guessing procedure is based on mathematical analysis to the parsimonious state space model and the statistical models for the inputs. In the following subsections, we firstly investigate the input series and the hidden states of the dynamics system in Sect. 4.5.1. Next, in Sect. 4.5.2, we model the relationship between the output series and the two lagged input series with linear regression, and the cross covariance between the output series and the two-input series. Then, in Sect. 4.5.3, we apply what we have obtained from Sects. 4.5.1 and 4.5.2 in guessing the unknown parameters. Finally, to avoid confusion caused by the difference between two procedures: one for the mathematical analysis and the other for the guessing of the unknown parameters, we explain the steps of the guessing procedure in Sect. 4.5.4.

4.5.1 Investigation on the input series and the hidden states

The investigation on the input series is the first step for the mathematical analysis. We assume that the input series for our model are orthogonal and autoregressive as shown in Eqs. 4. 5.1 and 4.5.2,

$$u_{1,t} = u_{1,b} + \rho_1 u_{1,t-1} + \zeta_{1,t}, \quad (4.5.1)$$

$$u_{2,t} = u_{2,b} + \rho_2 u_{2,t-1} + \zeta_{2,t}, \quad (4.5.2)$$

where $u_{1,b}$ and $u_{2,b}$ are constants, ρ_1 and ρ_2 are the 1st order autocorrelation coefficients, while $\zeta_{1,t}$ and $\zeta_{2,t}$ are Gaussian distributed stimuli with $\zeta_{1,t} \sim N(0, v_{\zeta_1})$ and $\zeta_{2,t} \sim N(0, v_{\zeta_2})$. The means of $u_{1,t}$ and $u_{2,t}$ are $u_{1,\infty} = \frac{u_{1,b}}{1-\rho_1}$, and $u_{2,\infty} = \frac{u_{2,b}}{1-\rho_2}$, respectively. Moreover, denoting $v_{1,t} = \sum_{i=0}^{\infty} \rho_1^i \zeta_{1,t-i}$ and $v_{2,t} = \sum_{i=0}^{\infty} \rho_2^i \zeta_{2,t-i}$, we have the variances (when the integer, $h = 0$) and the autocovariances (when $h > 0$) of $v_{1,t}$ and $v_{2,t}$:

$$\gamma_{v_1}^{(h)} = \rho_1^h v_{\zeta_1} / (1 - \rho_1^2), \quad (4.5.3)$$

$$\gamma_{v_2}^{(h)} = \rho_2^h v_{\zeta_2} / (1 - \rho_2^2). \quad (4.5.4)$$

On the other hand, from the parsimonious state space model, we have

$$x_{1,t} = k_1 (\gamma_1 u_{1,t-1} + \gamma_2 u_{2,t-1}) + \xi_t, \quad (4.5.5)$$

$$x_{2,t} = \gamma_1 u_{1,t} + \gamma_2 u_{2,t} + e_{2,t}, \quad (4.5.6)$$

where

$$k_1 = \beta / (1 - \alpha), \quad (4.5.7)$$

$$\xi_t = \sum_{i=0}^{\infty} \alpha^i (\beta e_{2,t-i-2} + e_{1,t-i-1}) \quad (4.5.8)$$

with ξ_t 's, or $x_{1,t}$'s variance and autocovariance,

$$v_{\xi} = (q_1 + 2\alpha\beta q_x + \beta^2 q_2) / (1 - \alpha^2), \quad (4.5.9)$$

$$\gamma_{\xi}^{(h)} = v_{\xi} \alpha^h, (h > 0). \quad (4.5.10)$$

Moreover, applying Eqs. 4.5.5 and 4.5.6, we obtain the covariance between $x_{1,t}$ and $x_{2,t}$ as 0 and the variance of $x_{2,t}$ as q_2 . Assuming the statistical indices of the hidden states, $x_{1,t}$ and $x_{2,t}$, are time-invariant, we can obtain the variances and covariance of the initial states from the variances of process noise series:

$$v_{10} = v_{\xi}, \quad (4.5.11)$$

$$v_{x0} = 0. \quad (4.5.12)$$

$$v_{20} = q_2 \quad (4.5.13)$$

By the way, from the parsimonious state space model, we notice that $\mu_{x_{1,t}} = k_1 \mu_{x_{2,t}}$. Considering that the state space model under investigation is Gaussian, it is reasonable to assume the variance, $v_{x_{1,t}}$, of $x_{1,t}$, and the variance, $v_{x_{2,t}}$, of $x_{2,t}$, satisfy $v_{x_{1,t}} = k_1^2 v_{x_{2,t}}$. Accordingly, we have

$$v_\xi = k_1^2 q_2. \quad (4.5.14)$$

Then we have a nonlinear constraint as below:

$$(q_1 + 2\alpha\beta q_x + \beta^2 q_2)/(1 - \alpha^2) = \beta^2 q_2/(1 - \alpha)^2. \quad (4.5.15)$$

Such a nonlinear constraint is only one example of various constraints to the parameters.

4.5.2 Linear regression and cross covariance between input

series and output series

In this subsection, we explore the linear regression of the output series with respect to the two lagged input series, as well as the cross covariances between the output series and the two input series.

Denote $\eta_t = \xi_t + e_{2,t}$, we have the variance and autocovariance of η_t

$$v_\eta = v_\xi + q_2, \quad (4.5.16)$$

$$\gamma_\eta^{(h)} = \gamma_\xi^{(h)} + \beta\alpha^{h-2}q_2 + \alpha^{h-1}q_x. \quad (4.5.17)$$

Denote

$$y_b = \gamma_1 u_{1,b} + \gamma_2 u_{2,b}, \quad (4.5.18)$$

$$\theta_1 = \gamma_1(k_1 + \rho_1), \quad (4.5.19)$$

$$\theta_2 = \gamma_2(k_1 + \rho_2), \quad (4.5.20)$$

$$\varepsilon_t = \gamma_1 \zeta_{1,t} + \gamma_2 \zeta_{2,t} + \eta_t + w_t, \quad (4.5.21)$$

we have y_t expressed in linear regression in (4.5.22) where y_b , θ_1 and θ_2 are the regression parameters to be estimated

$$y_t = y_b + \theta_1 u_{1,t-1} + \theta_2 u_{2,t-1} + \varepsilon_t \quad (4.5.22)$$

with the mean of y_t

$$\mu_y = (1 + k_1)(\gamma_1 u_{1,\infty} + \gamma_2 u_{2,\infty}) \quad (4.5.23)$$

and the variance and autocovariance of ε_t are

$$v_\varepsilon = \gamma_1^2 v_{\zeta_1} + \gamma_2^2 v_{\zeta_2} + \sigma_\eta^2 + r, \quad (4.5.24)$$

$$\gamma_\varepsilon^{(h)} = \gamma_\eta^{(h)}. \quad (4.5.25)$$

On the other hand, since the -1st order cross covariance between y_t and $u_{1,t}$ and $u_{2,t}$ are defined as

$$\gamma_{u_1 y}^{(-1)} = E[(u_{1,t-1} - u_{1,\infty})(y_{t-1} - \mu_y)],$$

and

$$\gamma_{u_2 y}^{(-1)} = E[(u_{2,t-1} - u_{2,\infty})(y_{t-1} - \mu_y)],$$

we have

$$\gamma_{u_1 y}^{(-1)} = E[v_{1,t-1}(\theta_1 v_{1,t-2} + \theta_2 v_{2,t-2} + \varepsilon_{t-1})], \quad (4.5.26)$$

$$\gamma_{u_2 y}^{(-1)} = E[v_{2,t-1}(\theta_1 v_{1,t-2} + \theta_2 v_{2,t-2} + \varepsilon_{t-1})]. \quad (4.5.27)$$

Then we can obtain Eqs. 4.5.28 and 4.5.29 about θ_1 , θ_2 , γ_1 and γ_2 :

$$\theta_1 \gamma_{v_1}^{(1)} + \gamma_1 v_{\zeta_1} = \gamma_{u_1 y}^{(-1)}, \quad (4.5.28)$$

$$\theta_2 \gamma_{v_2}^{(1)} + \gamma_2 v_{\zeta_2} = \gamma_{u_2 y}^{(-1)}. \quad (4.5.29)$$

Applying Eqs. 4.5.18, 4.5.19 and 4.5.20 to Eqs. 4.5.28 and 4.5.29, we have equations for k_1 , γ_1 and γ_2

$$\gamma_1(k_1 + \rho_1)\gamma_{v_1}^{(1)} + \gamma_2 v_{\zeta_1} = \gamma_{u_1 y}^{(-1)}, \quad (4.5.30)$$

$$\gamma_2(k_1 + \rho_2)\gamma_{v_2}^{(1)} + \gamma_2 v_{\zeta_2} = \gamma_{u_2 y}^{(-1)}. \quad (4.5.31)$$

Applying Eqs. 4.5.3 and 4.5.4, we can express γ_1 and γ_2 using k_1 ,

$$\gamma_1 = (\gamma_{u_1 y}^{(-1)} / v_{\zeta_1})(1 - \rho_1^2) / (1 + \rho_1 k_1), \quad (4.5.32)$$

$$\gamma_2 = (\gamma_{u_2 y}^{(-1)} / v_{\zeta_2})(1 - \rho_2^2) / (1 + \rho_2 k_1). \quad (4.5.33)$$

Assuming $\mu_y \neq 0$ and denoting

$$m_1 = (1 - \rho_1^2)(u_{1,\infty} / \mu_y)(\gamma_{u_1 y}^{(-1)} / v_{\zeta_1}), \quad (4.5.34)$$

$$m_2 = (1 - \rho_2^2)(u_{2,\infty} / \mu_y)(\gamma_{u_2 y}^{(-1)} / v_{\zeta_2}), \quad (4.5.35)$$

then applying Eqs. 4.5.32 and 4.5.33 to Eqs. 4.5.18, 4.5.19, 4.5.20 and 4.5.23, with some manipulation, we obtain

$$y_b / \mu_y = m_1(1 - \rho_1) / (1 + \rho_1 k_1) + m_2(1 - \rho_2) / (1 + \rho_2 k_1), \quad (4.5.36)$$

$$\theta_1 = m_1(\mu_y / u_{1,\infty})(k_1 + \rho_1) / (1 + \rho_1 k_1), \quad (4.5.37)$$

$$\theta_2 = m_2(\mu_y / u_{2,\infty})(k_1 + \rho_2) / (1 + \rho_2 k_1), \quad (4.5.38)$$

$$m_1 / (1 + \rho_1 k_1) + m_2 / (1 + \rho_2 k_1) = 1 / (1 + k_1). \quad (4.5.39)$$

Denote

$$a = \rho_2 m_1 + \rho_1 m_2 - \rho_1 \rho_2, \quad (4.5.40)$$

$$b = \rho_2 m_1 + \rho_1 m_2 + m_1 + m_2 - \rho_1 - \rho_2, \quad (4.5.41)$$

$$c = m_1 + m_2 - 1, \quad (4.5.42)$$

we can transform Eq. 4.5.39 into a quadratic equation:

$$a k_1^2 + b k_1 + c = 0. \quad (4.5.43)$$

From Eq. 4.1.3, we assume $\beta > 0$. Considering Eq. 4.5.7, we have $k_1 > 0$. As one of the roots of Eq. 4.5.43, we have

$$k_1 = (-b + \sqrt{b^2 - 4ac})/(2a). \quad (4.5.44)$$

After the value of k_1 value is available, applying Eqs. 4.5.36, 4.5.37 and 4.5.38, we can obtain the values of y_b , θ_1 and θ_2 , the parameters of the regression represented by Eq. 4.5.22. Furthermore, applying Eqs. 4.5.19 and 4.5.20, we obtain the parameters γ_1 and γ_2 :

$$\gamma_1 = \theta_1/(k_1 + \rho_1), \quad (4.5.45)$$

$$\gamma_2 = \theta_2/(k_1 + \rho_2). \quad (4.5.46)$$

4.5.3 Obtaining the guessed value of the unknown parameters

In this subsection, we apply what we have obtained from Sects. 4.5.1 and 4.5.2 in guessing the unknown parameters. From the parsimonious state space model, we have

$$\mu_y = (1 + k_1)(\gamma_1 u_{1,\infty} + \gamma_2 u_{2,\infty}). \quad (4.5.47)$$

Since assume $u_{1,t}$ and $u_{2,t}$ are orthogonal, the covariance between $v_{1,t}$ and $v_{2,t}$ is insignificant, considering Eqs. 4.5.21 and 4.5.22, then denoting

$$v_u = \theta_1^2 v_{v_1} + \gamma_1^2 v_{\zeta_1} + \theta_2^2 v_{v_1} + \gamma_2^2 v_{\zeta_2} + 2\theta_1\theta_2 v_{v_x}, \quad (4.5.48)$$

we obtain the variance and the autocovariance of y_t :

$$v_y = v_u + v_\xi + q_2 + r, \quad (4.5.49)$$

$$\gamma_y^{(h)} = \theta_1^2 \gamma_{v_1}^{(h)} + \theta_1 \gamma_1 \rho_1^{h-1} v_{\zeta_1} + \theta_2^2 \gamma_{v_2}^{(h)} + \theta_2 \gamma_2 \rho_2^{h-1} v_{\zeta_2} + \gamma_\varepsilon^{(h)}. \quad (4.5.50)$$

When the inputs $u_{1,t}$ and $u_{2,t}$ are neither autoregressive nor correlative, we have ρ_1 and ρ_2 equal to zero,

$$\gamma_y^{(1)} = \theta_1 \gamma_1 v_{\zeta_1} + \theta_2 \gamma_2 v_{\zeta_2} + \gamma_\varepsilon^{(1)}, \quad (4.5.51)$$

$$\gamma_y^{(h)} = \gamma_\varepsilon^{(h)} \quad (h > 1). \quad (4.5.52)$$

Applying Eqs. 4.5.10 and 4.5.17, we can guess the value of α from the autocovariance of the output series

$$\alpha = \gamma_y^{(3)} / \gamma_y^{(2)}. \quad (4.5.53)$$

Applying Eq. 4.5.7, we can further guess the value for β ,

$$\beta = k_1(1 - \alpha). \quad (4.5.54)$$

After the values for k_1 , y_b , θ_1 and θ_2 , α and β are obtained, we can further guess the variances for the process noises and the observation noise. Assuming $q_x = 0$, and $r = k_2^2 q_2$, then applying Eqs. 4.5.7, 4.5.14 and 4.5.49, we have:

$$q_1 = [(1 - \alpha^2)k_1^2 - \beta^2](v_y - v_u)/(k_1^2 + 1 + k_2^2), \quad (4.5.55)$$

$$q_2 = (v_y - v_u)/(k_1^2 + 1 + k_2^2), \quad (4.5.56)$$

$$r = k_2^2(v_y - v_u)/(k_1^2 + 1 + k_2^2). \quad (4.5.57)$$

In practice, the parameters k_1 and k_2 define the distribution of the overall output noises among the process and the observation.

On the other hand, using Eqs. 4.5.1 and 4.5.2, when $\rho_1 = 0$ and $\rho_2 = 0$, we can estimate $u_{1,0}$ and $u_{2,0}$ as $\hat{u}_{1,0} = u_{1b}$ and $\hat{u}_{2,0} = u_{2b}$. When $\rho_1 \neq 0$ and $\rho_2 \neq 0$, we can estimate $u_{1,0}$, $u_{2,0}$ as $\hat{u}_{1,0} = (u_{1,1} - u_{1b})/\rho_1$ and $\hat{u}_{2,0} = (u_{2,1} - u_{2b})/\rho_2$. Applying to the parsimonious state space model, we can estimate the mean of the initial state, $x_{2,0}$ as:

$$\mu_{2,0} = \hat{x}_{2,0} = \gamma_1 u_{1,0} + \gamma_2 u_{2,0}. \quad (4.5.58)$$

Moreover, we can further estimate the means of $x_{2,1}$, $x_{1,1}$ and $x_{1,0}$ as below:

$$\hat{x}_{2,1} = \gamma_1 u_{1,1} + \gamma_2 u_{2,1}, \quad (4.5.59)$$

$$\hat{x}_{1,1} = y_1 - \hat{x}_{2,1}, \quad (4.5.60)$$

$$\mu_{1,0} = \hat{x}_{1,0} = (\hat{x}_{1,1} - \beta \hat{x}_{2,0})/\alpha. \quad (4.5.61)$$

4.5.4 Steps for the guessing

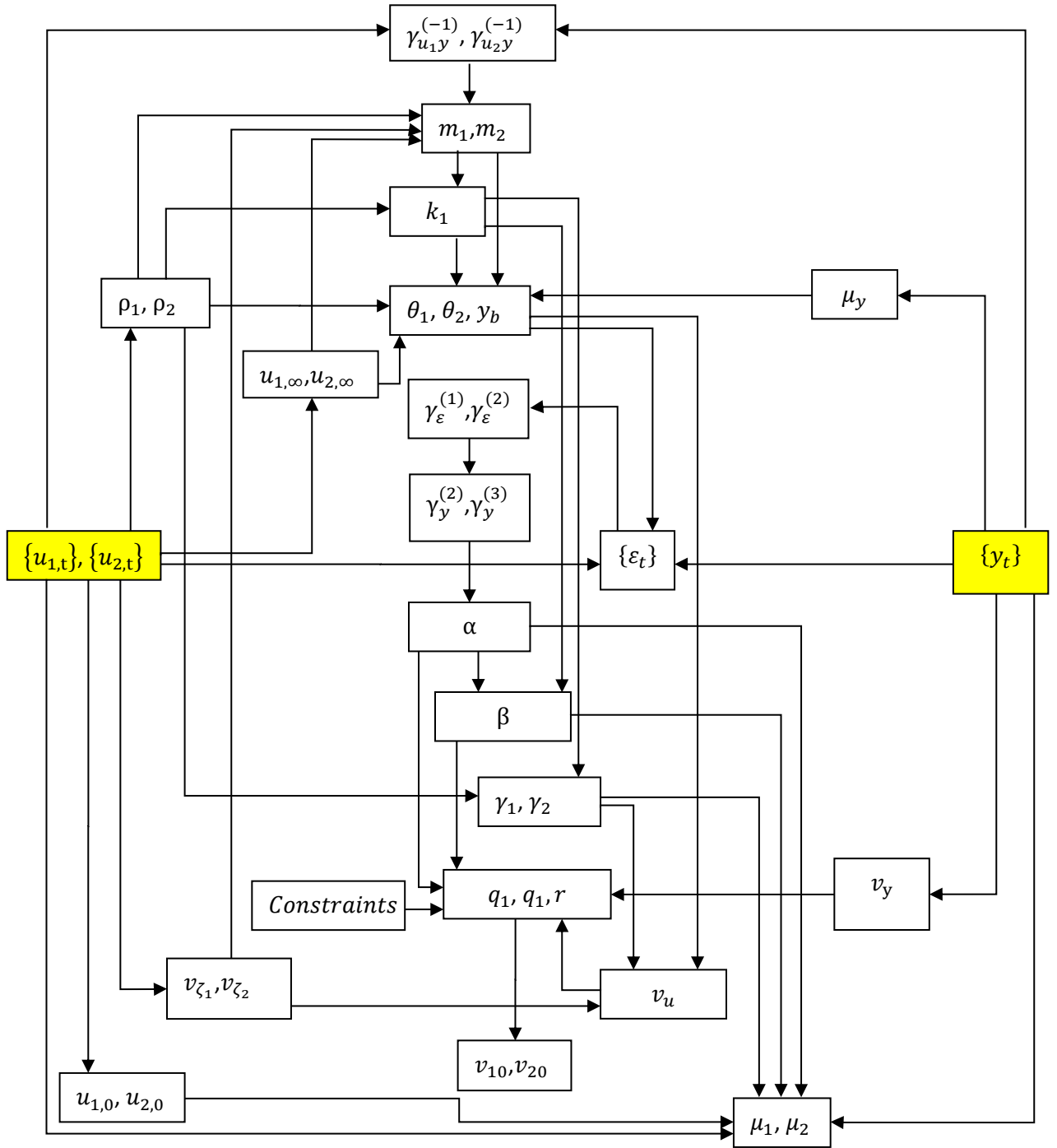


Figure 4.1: Directed Acyclic Graph for the Guessing Steps

Figure 4.1 illustrates the guessing procedure proposed by us. Assuming we have the input series, $\{u_{1,t}\}$ and $\{u_{2,t}\}$, and the output series $\{y_t\}$, the procedure starts from the calculation of the statistical indices, including $u_{1,\infty}$, $u_{2,\infty}$, ρ_1 , ρ_2 , and v_{ζ_1}, v_{ζ_2} of $\{u_{1,t}\}$ and $\{u_{2,t}\}$, and μ_y and v_y of $\{y_t\}$, which are used in the following major steps. Then, we investigate the statistical indices of regression error series, $\{\varepsilon_t\}$, from Eq. 4.5.21, obtain $\gamma_y^{(2)}$ and $\gamma_y^{(3)}$ from $\gamma_\varepsilon^{(2)}$ and $\gamma_\varepsilon^{(3)}$, applying Eq. 4.5.52, and calculate the transfer parameters, α and then β , using Eq. 4.5.53 and 4.5.54 respectively.

Furthermore, we apply constraints to obtain q_1, q_1, r , the parameters for the process noises and observation noises, using Eqs. 4.57 - 4.59, and calculate the variances and covariance, v_{10} , v_{x0} and v_{20} , of the initial state, using the Eqs. 4.5.9 and 4.5.11 - 4.5.13.

Finally, the means of initial inputs, $u_{1,0}, u_{2,0}$, are estimated using Eqs. 4.5.1 and 4.5.2. Then the means, μ_1 and μ_2 , of the initial states are calculated, using Eqs. 4.5.58 - 4.5.61.

Moreover, we also have proved the ergodicity of the state series, the output series, the noise series and the error series. As a result, we are actually using the time average of a series to represent the ensemble average of the same series. However, the proof about ergodicity is not provided so as to prevent this chapter from being lengthy.

4.6 MLE Using the EM Algorithm

In this section, we present how MLE is implemented through the EM algorithm to find the values of the unknown parameters.

As we know from Eqs. 4.4.1 - 4.4.3, the mathematical optimization problem for our problem is different from the mathematical optimization problem shown in Eqs. 4.2.12-

4.2.14. In this section, we consider two constraints represented by Eq. 4.5.15 and Eq. 4.5.49, denoting $v_n = \sigma_y^2 - \sigma_u^2$:

$$h_1(\boldsymbol{\psi}) = (1 - \alpha)q_1 + 2\alpha(1 - \alpha)\beta q_x - 2\alpha\beta^2 q_2, \quad (4.6.1)$$

$$h_2(\boldsymbol{\psi}) = q_1 + 2\alpha\beta q_x + (1 - \alpha^2 + \beta^2)q_2 + (1 - \alpha^2)(r - v_n). \quad (4.6.2)$$

The mathematical optimization problem formulated is:

$$\min_{\boldsymbol{\psi}} D_y(\boldsymbol{\psi}) \quad (4.6.3)$$

subjected to

$$\begin{cases} h_1(\boldsymbol{\psi}) = 0 \\ h_2(\boldsymbol{\psi}) = 0 \end{cases} \quad (4.6.4)$$

Applying the method of Lagrange multipliers, we obtain the Lagrange function as a new objective function:

$$L_y(\boldsymbol{\psi}, \lambda_1, \lambda_2) = D_y(\boldsymbol{\psi}) + \lambda_1 h_1(\boldsymbol{\psi}) + \lambda_2 h_2(\boldsymbol{\psi}), \quad (4.6.5)$$

where λ_1 and λ_2 are Lagrange multipliers. And the optimization problem become:

$$\min_{\boldsymbol{\psi}, \lambda_1, \lambda_2} L_y(\boldsymbol{\psi}, \lambda_1, \lambda_2). \quad (4.6.6)$$

Per Eqs. 4.5.11 - 4.5.13, the parameters, v_{10}, v_{x0}, v_{20} , about the variances and covariance of the initial hidden state, can be estimated from parameters, q_1, q_x, q_2 , about the variances and covariance of the process noises. As a result, the parameter vector $\boldsymbol{\psi}_{13}$ can be replaced by a shorter vector $\boldsymbol{\psi}_{10}$:

$$\boldsymbol{\psi}_{10} = [\alpha, \beta, \gamma_1, \gamma_2, q_1, q_x, q_2, r, \mu_{10}, \mu_{20}]. \quad (4.6.7)$$

In Eqs. 4.6.1-4.6.2, $\boldsymbol{\psi} = \boldsymbol{\psi}_{10}$.

In the k th iteration of the E step of the EM algorithm, using the parameters estimated in the $(k-1)$ th iteration, we perform the Kalman filtering and the Kalman smoothing to estimate the means and variances about the hidden states, and calculate the conditional

expectation of the log likelihood based on complete data. For the first iteration, we use the values of parameters guessed in Sect. 4.5.3, as the starting values for the parameters:

$$\boldsymbol{\Psi}_{10}(0) = [\alpha(0), \beta(0), \gamma_1(0), \gamma_2(0), q_1(0), q_x(0), q_2(0), r(0), \mu_1(0), \mu_2(0)]. \quad (4.6.8)$$

For the Lagrange multipliers, we assign them as zeros. During the M step of the EM iteration, we estimate the values of the unknown parameters,

$$\boldsymbol{\Psi}_{10}(k) = [\alpha(k), \beta(k), \gamma_1(k), \gamma_2(k), q_1(k), q_x(k), q_2(k), r(k), \mu_1(k), \mu_2(k)], \quad (4.6.9)$$

and the Lagrange multipliers, $\boldsymbol{\lambda}(k) = \{\lambda_1(k), \lambda_2(k)\}$, through the maximization of the conditional expectation of the log likelihood calculated using $\boldsymbol{\Psi}_{10}(k-1)$ and $\boldsymbol{\lambda}(k)$, on the E step.

In the M step, from the first order necessary conditions, we apply certain features of the model and constraints. We directly estimate the values of four parameters first, then solve other simultaneous nonlinear equations obtained, employing a trust region algorithm.

Accordingly, in this section, in sequence, we introduce the construction of the expected value of the objective function based on complete data, for the E step, and then the first order necessary conditions and the trust region method for the M step.

4.6.1 “Expectation” value for the E step

In the E step, we will assume the parameters are known. For convenience of notation, we dropped k in the following sections. After the Kalman filtering and the Kalman smothering have been implemented according to Sect. 4.2.1, we obtained the state, \mathbf{x}_t . Then the conditional densities for output, y_t , and the state, \mathbf{x}_t , are

$$P(y_t|\mathbf{x}_t) = (2\pi|\mathbf{R}|)^{-\frac{1}{2}} \exp[-(y_t - \mathbf{H}\mathbf{x}_t)^2(2\mathbf{R})^{-1}], \quad (4.6.10)$$

$$P(\mathbf{x}_t|\mathbf{x}_{t-1}) = (2\pi)^{-1}|\mathbf{Q}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\mathbf{x}_t - \mathbf{\Phi}\mathbf{x}_{t-1} - \mathbf{\Gamma}\mathbf{u}_t)'\mathbf{Q}^{-1}(\mathbf{x}_t - \mathbf{\Phi}\mathbf{x}_{t-1} - \mathbf{\Gamma}\mathbf{u}_t)\right]. \quad (4.6.11)$$

Assuming a Gaussian initial state density, we have

$$P(\mathbf{x}_0) = (2\pi)^{-1}|\mathbf{Q}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\mathbf{x}_0 - \mathbf{\mu}_0)'\mathbf{Q}^{-1}(\mathbf{x}_0 - \mathbf{\mu}_0)\right]. \quad (4.6.12)$$

By the Markov property implicated in this state space model,

$$P(\{\mathbf{x}\}, \{\mathbf{y}\}) = P(\mathbf{x}_0) \prod_{t=1}^T P(\mathbf{x}_t|\mathbf{x}_{t-1}) \prod_{t=1}^T P(\mathbf{y}_t|\mathbf{x}_t). \quad (4.6.13)$$

Denoting the expected objection function value based on complete data $\{\mathbf{x}\}$ and $\{\mathbf{y}\}$

$$\Omega(\mathbf{\Psi}) = -\ln P(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}) + \sum_{i=1}^S \lambda_i h_i(\mathbf{\Psi}), \quad (4.6.14)$$

and

$$\Omega_0(\mathbf{\Psi}) = (\mathbf{x}_0 - \mathbf{\mu}_0)'\mathbf{Q}^{-1}(\mathbf{x}_0 - \mathbf{\mu}_0) + \ln|\mathbf{Q}| + \ln 2\pi, \quad (4.6.15)$$

$$\Omega_1(\mathbf{\Psi}) = \sum_{t=1}^T (\mathbf{x}_t - \mathbf{\Phi}\mathbf{x}_{t-1} - \mathbf{\Gamma}\mathbf{u}_t)'\mathbf{Q}^{-1}(\mathbf{x}_t - \mathbf{\Phi}\mathbf{x}_{t-1} - \mathbf{\Gamma}\mathbf{u}_t) + T\ln|\mathbf{Q}| + T\ln 2\pi, \quad (4.6.16)$$

$$\Omega_2(\mathbf{\Psi}) = \sum_{t=1}^T (\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)^2/\mathbf{R} + T\ln|\mathbf{R}| + T\ln 2\pi, \quad (4.6.17)$$

we have,

$$\Omega = \frac{1}{2}\Omega_0 + \frac{1}{2}\Omega_1 + \frac{1}{2}\Omega_2 + \sum_{i=1}^S \lambda_i h_i(\mathbf{\Psi}). \quad (4.6.18)$$

After we obtain the expected values for \mathbf{x}_0 and \mathbf{x}_t as $\mathbf{x}_0^T \equiv E[\mathbf{x}_0|\{\mathbf{y}, \mathbf{u}\}]$ and $\mathbf{x}_t^T \equiv E[\mathbf{x}_t|\{\mathbf{y}, \mathbf{u}\}]$ respectively, we can calculate the “expectation” of the objective Ω for the k th iteration:

$$E(\Omega) = \frac{1}{2}E(\Omega_0) + \frac{1}{2}E(\Omega_1) + \frac{1}{2}E(\Omega_2) + \sum_{i=1}^S \lambda_i h_i(\mathbf{\Psi}). \quad (4.6.19)$$

As we have presented in Sect. 4.4, some elements of the matrices for the dynamic systems are already known, thus we investigate the expected value of the objective function

with respect to unknown elements. Conversely, the expected value of the objective function is traditionally investigated with respect to unknown matrices directly. This is one of the significant differences in our proposal to implement the EM algorithm. As a result, if it is possible, we attempt to find the scalar expression for the expected value of the objective function. In this section, we will see how far we can go to reach the goal.

Denoting $\overline{x_{1,0}^2} = E(x_{1,0}^2)$, $\overline{x_{2,0}^2} = E(x_{2,0}^2)$, $\overline{x_{1,0}x_{2,0}} = E(x_{1,0}x_{2,0})$, $\overline{x_{1,0}} = E(x_{1,0})$, and $\overline{x_{2,0}} = E(x_{2,0})$, we have,

$$E(\Omega_0) = (q_1q_2 - q_x^2)^{-1} [q_2(\overline{x_{1,0}^2} - 2\mu_{10}\overline{x_{1,0}} + \mu_{10}^2) - 2q_x(\overline{x_{1,0}x_{2,0}} - \mu_{20}\overline{x_{1,0}} - \mu_{10}\overline{x_{2,0}} + \mu_{10}\mu_{20}) + q_1(\overline{\mu_{20}^2} - 2\mu_{20}\overline{x_{2,0}} + \mu_{20}^2)] + \ln(q_1q_2 - q_x^2). \quad (4.6.20)$$

Further, we denote

$$P_{\sigma 0} = \overline{x_{1,0}^2} - 2\mu_{10}\overline{x_{1,0}} + \mu_{10}^2, \quad (4.6.21)$$

$$M_{\sigma 0} = \overline{x_{1,0}x_{2,0}} - \mu_{20}\overline{x_{1,0}} - \mu_{10}\overline{x_{2,0}} + \mu_{10}\mu_{20}, \quad (4.6.22)$$

$$U_{\sigma 0} = \overline{x_{2,0}^2} - 2\mu_{20}\overline{x_{2,0}} + \mu_{20}^2. \quad (4.6.23)$$

Then we obtain,

$$E(\Omega_0) = (q_1q_2 - q_x^2)^{-1} (q_2P_{\sigma 0} - 2q_xM_{\sigma 0} + q_1U_{\sigma 0}) + \ln(q_1q_2 - q_x^2) + \ln 2\pi. \quad (4.6.24)$$

Furthermore, apply Eq. 4.6.16 using the matrices for the parsimonious state space model, we arrive at,

$$E(\Omega_1) = (q_1q_2 - q_x^2)^{-1} E \left[\sum_{t=1}^T \left\{ q_2 [x_{1,t} - (\alpha x_{1,t-1} + \beta x_{2,t-1})]^2 - 2q_x [x_{2,t} - (\gamma_1 u_{1,t} + \gamma_2 u_{2,t})] [x_{1,t} - (\alpha x_{1,t-1} + \beta x_{2,t-1})] + q_1 [x_{2,t} - (\gamma_1 u_{1,t} + \gamma_2 u_{2,t})]^2 \right\} \right] + T \ln(q_1q_2 - q_x^2) + T \ln 2\pi. \quad (4.6.25)$$

Denoting

$$\begin{cases} P_{11} = \sum_{t=1}^T E(x_{1,t}^2) \\ P_{12} = \sum_{t=1}^T E(x_{1,t}x_{2,t}) \\ P_{22} = \sum_{t=1}^T E(x_{2,t}^2) \\ P_{xx} = \sum_{t=1}^T E(x_{1,t-1}^2) \\ P_{yy} = \sum_{t=1}^T E(x_{2,t-1}^2) \end{cases},$$

$$\begin{cases} P_{xy} = \sum_{t=1}^T E(x_{1,t-1}x_{2,t-1}) \\ P_{x1} = \sum_{t=1}^T E(x_{1,t-1}x_{1,t}) \\ P_{x2} = \sum_{t=1}^T E(x_{1,t-1}x_{2,t}) \\ P_{y1} = \sum_{t=1}^T E(x_{2,t-1}x_{1,t}) \\ P_{y2} = \sum_{t=1}^T E(x_{2,t-1}x_{2,t}) \end{cases},$$

$$\begin{cases} M_{11} = \sum_{t=1}^T E(x_{1,t}u_{1,t}) \\ M_{12} = \sum_{t=1}^T E(x_{1,t}u_{2,t}) \\ M_{21} = \sum_{t=1}^T E(x_{2,t}u_{1,t}) \\ M_{22} = \sum_{t=1}^T E(x_{2,t}u_{2,t}) \end{cases},$$

$$\begin{cases} M_{x1} = \sum_{t=1}^T E(x_{1,t-1}u_{1,t}) \\ M_{x2} = \sum_{t=1}^T E(x_{1,t-1}u_{2,t}) \\ M_{y1} = \sum_{t=1}^T E(x_{2,t-1}u_{1,t}) \\ M_{y2} = \sum_{t=1}^T E(x_{2,t-1}u_{2,t}) \end{cases},$$

and

$$\begin{cases} U_{11} = \sum_{t=1}^T (u_{1,t}^2) \\ U_{22} = \sum_{t=1}^T (u_{2,t}^2) \\ U_{12} = \sum_{t=1}^T (u_{1,t}u_{2,t}) \end{cases},$$

we obtain

$$\begin{aligned} E(\Omega_1) = & (q_1q_2 - q_x^2)^{-1} \{ q_2 [-(\alpha P_{x1} + \beta P_{y1} - P_{11}) + \alpha(\alpha P_{xx} + \beta P_{xy} - P_{x1}) + \\ & \beta(\beta P_{yy} + \alpha P_{xy} - P_{y1})] - 2q_x [-(\gamma_1 M_{11} + \gamma_2 M_{12} - P_{12}) + \alpha(\gamma_1 M_{x1} + \gamma_2 M_{x2} - P_{x2}) + \\ & \beta(\gamma_1 M_{y1} + \gamma_2 M_{y2} - P_{y2})] + q_1 [-(\gamma_1 M_{21} + \gamma_2 M_{22} - P_{22}) + \gamma_1(\gamma_1 U_1 + \gamma_2 U_x - M_{21}) + \\ & \gamma_2(\gamma_1 U_x + \gamma_2 U_2 - M_{22})] \} + T \ln(q_1q_2 - q_x^2) + \ln 2\pi. \end{aligned} \quad (4.6.26)$$

In addition, we denote

$$P_\alpha = \alpha P_{xx} + \beta P_{xy} - P_{x1}, \quad (4.6.27)$$

$$P_\beta = \alpha P_{xy} + \beta P_{yy} - P_{y1}, \quad (4.6.28)$$

$$P_1 = \alpha P_{x1} + \beta P_{y1} - P_{11}, \quad (4.6.29)$$

$$M_\alpha = \gamma_1 M_{x1} + \gamma_2 M_{x2} - P_{x2}, \quad (4.6.30)$$

$$M_\beta = \gamma_1 M_{y1} + \gamma_2 M_{y2} - P_{y2}, \quad (4.6.31)$$

$$M_1 = r_1 M_{11} + r_2 M_{12} - P_{12}, \quad (4.6.32)$$

$$U_{\gamma_1} = \gamma_1 U_{11} + \gamma_2 U_{12} - M_{21}, \quad (4.6.33)$$

$$U_{\gamma_2} = \gamma_2 U_{22} + \gamma_1 U_{12} - M_{22}, \quad (4.6.34)$$

$$U_1 = \gamma_1 M_{21} + \gamma_2 M_{22} - P_{22}. \quad (4.6.35)$$

Then we simplify the expression for $E(\Omega_1)$ as,

$$\begin{aligned} E(\Omega_1) = & (q_1 q_2 - q_x^2)^{-1} \{ q_2 [\alpha P_\alpha + \beta P_\beta - P_1] - 2q_x (\alpha M_\alpha + \beta M_\beta - M_1) + \\ & q_1 (\gamma_1 U_{\gamma_1} + \gamma_2 U_{\gamma_2} - U_1) \} + T \ln(q_1 q_2 - q_x^2) + T \ln 2\pi. \end{aligned} \quad (4.6.37)$$

Further, denoting

$$P_\sigma = \alpha P_\alpha + \beta P_\beta - P_1, \quad (4.6.38)$$

$$M_\sigma = \alpha M_\alpha + \beta M_\beta - M_1, \quad (4.6.39)$$

$$U_\sigma = \gamma_1 U_{\gamma_1} + \gamma_2 U_{\gamma_2} - U_1, \quad (4.6.40)$$

we obtain a simplified expression for (Ω_1) :

$$\begin{aligned} E(\Omega_1) = & (q_1 q_2 - q_x^2)^{-1} \{ q_2 P_\sigma - 2q_x M_\sigma + q_1 U_\sigma \} + T \ln(q_1 q_2 - q_x^2) + \\ & T \ln 2\pi. \end{aligned} \quad (4.6.41)$$

Moreover, denoting $Y = \sum_{t=1}^T y_t^2$, $W_1 = \sum_{t=1}^T x_{1,t} y_t$, $W_2 = \sum_{t=1}^T x_{2,t} y_t$, and

$$P_w = Y + P_{11} + P_{22} - 2W_1 - 2W_2 + 2P_{12}, \quad (4.6.42)$$

we obtain

$$E(\Omega_2) = r^{-1}P_w + T \log r + T \ln 2\pi. \quad (4.6.43)$$

4.6.2 Mathematical optimization problem for the M step

In the M step, we assume that the parameters are unknown. Accordingly, $P_\alpha, P_\beta, P_1, M_\alpha, M_\beta, M_1, M_{\gamma_1}, M_{\gamma_2}, U_{\gamma_1}, U_{\gamma_2}, U_1, P_{\sigma 0}, M_{\sigma 0}, U_{\sigma 0}, P_\sigma, M_\sigma, U_\sigma, E(\Omega_0), E(\Omega_1), E(\Omega_2)$ and $E(\Omega)$ are all the functions with respect to the unknown parameters. We attempt to identify the estimates of these unknown parameters, for the k step of the EM iteration. The estimates are obtained when the expected value of the objective function, $E(\Omega)$, reaches its maximum with these estimates, subjected to two constraints represented by Eq. 4.6.1 and Eq. 4.6.2. The optimization problem is:

The mathematical optimization problem formulated is:

$$\min_{\psi} E(\Omega(\psi)) \quad (4.6.44)$$

where the objective function is

$$\Omega(\psi) = \frac{1}{2}E(\Omega_0(\psi)) + \frac{1}{2}E(\Omega_1(\psi)) + \frac{1}{2}E(\Omega_2(\psi)) + \sum_{i=1}^s \lambda_i h_i(\psi). \quad (4.6.45)$$

4.6.3 Using first order necessary conditions to estimate μ_{10} ,

μ_{20}, γ_1 and γ_2

We notice that some unknown parameters appear only in the expression of (Ω_0) , $E(\Omega_1)$ or $E(\Omega_2)$. For the parsimonious state space model, the means, μ_{10} and μ_{20} , of initial states, are for $E(\Omega_0)$ only, while the unknown elements, γ_1 and γ_2 , of the input matrix, \mathbf{F} , are for $E(\Omega_1)$ only. Moreover, all these four parameters are not constrained. With these considerations, we may apply first order sufficient conditions directly to obtain the estimates of these parameters.

Applying first order sufficient conditions to $E(\Omega_0)$ with respect to μ_{10} and μ_{20} and considering that $q_1 q_2 - q_x^2 > 0$, we have

$$\mu_{10} = \overline{x_{1,0}}, \quad (4.6.46)$$

$$\mu_{20} = \overline{x_{2,0}}. \quad (4.6.47)$$

Applying first order sufficient conditions to $E(\Omega_1)$ with respect to γ_1 and γ_2 and considering Eqs. 4.6.33 and 4.6.34, we can obtain γ_1 and γ_2 by solving simultaneous linear equations as below:

$$\gamma_1 U_{11} + \gamma_2 U_{12} = M_{21}, \quad (4.6.48)$$

$$\gamma_2 U_{22} + \gamma_1 U_{12} = M_{22}. \quad (4.6.49)$$

Therefore, we can estimate the values of the four parameters, μ_{10} and μ_{20} , and γ_1 and γ_2 , in the parsimonious state space model. By doing this, we can reduce the number of the unknown parameters to six before the formation of the nonlinear optimization problem with nonlinear constraints.

4.6.4 Nonlinear optimization methods

In Sect. 4.4, we concentrate on solving the mathematical optimization problem with respect to a set of only six unknown parameters: $\theta = [\alpha, \beta, q_1, q_x, q_2, r]$ and the Lagrange multipliers.

The M step of the EM algorithm is not completed although the values of μ_{10} , μ_{20} , γ_1 and γ_2 have been estimated. We continue with the attempt to find the optimal value of six unknown parameters with the objective function represented by Eq. 4.6.19 and two constraints represented by Eq. 4.6.1 and Eq. 4.6.2:

The mathematical optimization problem formulated for the M step is:

$$\min_{\boldsymbol{\theta}} E(\Omega) \quad (4.6.50)$$

subjected to Eqs. 4.6.4. Applying the method of Lagrange multipliers, the Lagrange function becomes:

$$L_{x,y}(\boldsymbol{\theta}, \lambda_1, \lambda_2) = E(\Omega_0(\boldsymbol{\theta})) + E(\Omega_1(\boldsymbol{\theta})) + E(\Omega_1(\boldsymbol{\theta})) + \lambda_1 h_1(\boldsymbol{\theta}) + \lambda_2 h_2(\boldsymbol{\theta}), \quad (4.6.51)$$

and the optimization problem becomes:

$$\min_{\boldsymbol{\theta}, \lambda_1, \lambda_2} L_{x,y}(\boldsymbol{\theta}, \lambda_1, \lambda_2), \quad (4.6.52)$$

where λ_1 and λ_2 are Lagrange multipliers.

Applying first order sufficient conditions to $L(\boldsymbol{\theta}, \lambda_1, \lambda_2)$ with respect to $\boldsymbol{\theta}, \lambda_1, \lambda_2$, we have below seven simultaneous nonlinear Eqs 4.6.53 - 4.6.59:

$$(q_2 P_\alpha - q_x M_\alpha) - \lambda_1 [0.5 q_1 - (1 - 2\alpha)\beta q_x + \beta^2 q_2] (q_1 q_2 - q_x^2) + \lambda_2 [\beta q_x - \alpha(q_2 + r - v_n)] (q_1 q_2 - q_x^2) = 0, \quad (4.6.53)$$

$$(q_2 P_\beta - q_x M_\beta) + \lambda_1 [\alpha(1 - \alpha)q_x - 2\alpha\beta q_2] (q_1 q_2 - q_x^2) + \lambda_2 (\alpha q_x + \beta q_2) (q_1 q_2 - q_x^2) = 0, \quad (4.6.54)$$

$$[(P_{\sigma 0} - q_1 + P_\sigma - T q_1) q_2^2 - 2 q_x q_2 (M_{\sigma 0} + M_\sigma) + q_x^2 (U_{\sigma 0} + q_2 + U_\sigma + T q_2)] - \lambda_1 (1 - \alpha) (q_1 q_2 - q_x^2)^2 - \lambda_2 (q_1 q_2 - q_x^2)^2 = 0, \quad (4.6.55)$$

$$[q_1 q_2 (M_{\sigma 0} + q_x + M_\sigma + T q_x) - q_x (q_2 P_{\sigma 0} + q_1 U_{\sigma 0} + q_2 P_\sigma + q_1 U_\sigma) + q_x^2 (M_{\sigma 0} - q_x + M_\sigma - T q_x)] - 2\lambda_1 \alpha (1 - \alpha) \beta (q_1 q_2 - q_x^2)^2 - 2\lambda_2 \alpha \beta (q_1 q_2 - q_x^2)^2 = 0, \quad (4.6.56)$$

$$[q_1^2 (U_{\sigma 0} - q_2 + U_\sigma - T q_2) - 2 q_1 q_x (M_{\sigma 0} + M_\sigma) + (P_{\sigma 0} + q_1 + P_\sigma + T q_1) q_x^2] + 2\lambda_1 \alpha \beta^2 (q_1 q_2 - q_x^2)^2 - \lambda_2 (1 - \alpha^2 + \beta^2) (q_1 q_2 - q_x^2)^2 = 0, \quad (4.6.57)$$

$$(1 - \alpha) q_1 + 2\alpha(1 - \alpha)\beta q_x - 2\alpha\beta^2 q_2 = 0, \quad (4.6.58)$$

$$q_1 + 2\alpha\beta q_x + (1 - \alpha^2 + \beta^2)q_2 + (1 - \alpha^2)(r - v_n) = 0. \quad (4.6.59)$$

To solve these simultaneous nonlinear equations, line search methods and trust region methods are two categories of methods employed frequently. A line search method first finds a descent direction and then computes a step size. The descent direction can be computed by various methods, such as gradient descent, Newton's method and Quasi-Newton method. Unlike line search methods, trust region methods compute a trial step by solving a trust region subproblem where a model function is minimized within a trust region. Accordingly, trust region methods improve robustness when starting far from the solution and handles ill-conditioned problems when Jacobian matrix used in line search is singular [135]. To further reduce the time spent in applying the trust region methods, approximation and heuristic strategies are used to restrict the trust-region subproblem to a two-dimensional subspace. Consequently, trust-region-reflective algorithm [136] is chosen to solve the simultaneous nonlinear equations in this chapter.

4.7 Simulation, Results and Discussion

The simulation to evaluate our proposed approach includes three major steps. Firstly, we preset exact values for the statistical indices of the inputs series and the exact values for parameters of the parsimonious state space model, generate input series, and then generate output series using Eqs. 4.1 and 4.2. The sample size is 1000. Secondly we guess the starting values of the parameters, applying the procedure introduced in Sect. 4.4. Finally, we implement the MLE method and the EM algorithm where we apply the Kalman filtering and the Kalman smoothing during the E step to calculate the conditionally expected log likelihood based on complete data and then apply the Lagrangian multiplier

method and trust-region-reflective algorithm during the M step to obtain the new parameter values from the simultaneous nonlinear equations: Eqs. 4.6.53 - 4.6.59.

Table 4.3: Statistical indices for input series

Sample Size	u_{1b}	ρ_1	v_{ζ_1}	u_{2b}	ρ_2	v_{ζ_2}
Exact	0	0	0.25	10	0	0.25
Guessed	0.014	-0.023	0.248	9.989	0.003	0.238

The exact values, the guessed values and the estimated values of the parameters are displayed in Table 4.3. In order to assess the performance of the guessing procedure and the estimation methods, root-mean-square deviation (RMSD) is used as a measure for the differences between the guessed/estimated parameter values and the exact parameter values. Considering that the values of different parameters are of different scales, we calculate the percentages of changes of the guessed/estimated values with respect to those of the exact values. The percentages are treated as a certain degree of normalization and are used to derive another metric named normalized root-mean-square deviation (NRMSD).

4.7.1 Generating input series and output series

We assume that the input series, $u_1(t)$ and $u_2(t)$, have no autocorrelation respectively but are orthogonal to each other.

Then we preset statistical indices for $u_1(t)$ and $u_2(t)$ as shown in the second row of Table 4.3. The input series are generated directly using these statistical indices.

To generate the output series, we need the “preset” values of parameters, y_b , θ_1 , θ_2 , v_ε , as shown in the second row of Table 4.4, and the preset values for the elements of Ψ_{10} , except q_x , as displayed in the second rows of Table 4.5 and Table 4.6 respectively. These

values of the regression parameters are treated as “preset” because they are calculated using the exact statistical indices of the input series, and the exact values of the parameters, Ψ_{10} , applying Eqs. 4.5.18 - 4.5.20. With all these parameters available and the input series generated, applying Eqs. 4.1 and 4.2, we can obtain the state series for hidden variables x_{1t} and x_{2t} , and the output series for the observation y_t .

Table 4.4: Regression parameters for the two-input model

Parameters	y_b	θ_1	θ_2	v_ε
“Preset”	12	2.4	1.92	2.168
Guessed	11.064	2.639	2.023	3.143

4.7.2 Guessing the parameters

When we are guessing the values of the parameters, we assume that we do not have the knowledge of the preset values of the parameters from Sect. 4.1. So firstly we must find the statistical indices of the input series directly from the generated input series for $u_1(t)$ and $u_2(t)$, which are shown in the third row of Table 4.3.

Then a critical step is to guess the values of parameters, y_b , θ_1 , θ_2 , v_ε , for the linear regression between the output series, $\{y_t\}$, and the lagged input series, $\{u_{1,t-1}\}$ and $\{u_{2,t-1}\}$, taking the advantage of cross variance between output series and input series. These parameters guessed using Eqs. 4.5.36 - 4.5.38 and 4.5.24 are displayed in the third row of Table 4.4.

All the values of the parameters of the parsimonious state space model can be guessed, continuing to follow the procedure illustrated in Figure 4.1.

The parameter values guessed are displayed in the third rows of Table 4.5.

4.7.3 Estimating the parameters

The guessed parameter values are used as starting values for the iteration of the EM algorithm. In the meantime, since $q_x = 0$, the first order sufficient conditions, i.e. Eqs. 4.6.53 - 4.6.59, can be further simplified as:

$$P_\alpha - [\lambda_1(0.5q_1 + \beta^2q_2) + \lambda_2\alpha(q_2 + r - v_n)]q_1 = 0 \quad (4.6.60)$$

$$P_\beta + (\lambda_2 - 2\lambda_1\alpha)\beta q_1q_2 = 0 \quad (4.6.61)$$

$$P_\sigma + P_{\sigma 0} - (T + 1)q_1 - [\lambda_1(1 - \alpha) + \lambda_2]q_1^2 = 0 \quad (4.6.62)$$

$$U_\sigma + U_{\sigma 0} - (T + 1)q_2 + [2\lambda_1\alpha\beta^2 - \lambda_2(1 - \alpha^2 + \beta^2)]q_2^2 = 0 \quad (4.6.63)$$

$$P_w - Tr - \lambda_2(1 - \alpha^2)r^2 = 0 \quad (4.6.64)$$

$$(1 - \alpha)q_1 - 2\alpha\beta^2q_2 = 0 \quad (4.6.65)$$

$$q_1 + (1 - \alpha^2 + \beta^2)q_2 + (1 - \alpha^2)(r - v_n) = 0 \quad (4.6.66)$$

Table 4.5: Comparison among preset/guessed and estimated parameters

Parameters	α	β	γ_1	γ_2	q_1	q_2	r	μ_1	μ_2	RMSE ₉	RMSE ₄
Exact	0.5	0.8	1.5	1.2	0.32	0.25	0.64	19.2	12	0	0
Guessed	0.62	0.70	1.44	1.09	0.45	0.27	0.70	21.82	10.92	0.90	0.10
Change (%)	24.6	-12.3	-4.2	-9.1	39.2	9.4	9.4	13.7	-9.0	17.75	14.63
Estimated	0.54	0.71	1.46	1.22	0.43	0.31	0.58	22.16	11.09	0.98	0.055
Change (%)	8.52	-11.34	-2.69	1.89	34.58	22.80	-8.96	15.43	-7.57	16.00	7.28

Assuming the starting value of the Lagrangian multiplier $\lambda=0$, Using the guessed values of the parameters and applying the trust-region-reflective algorithm, we obtain the estimates of the parameters, which are listed in the fifth row of Table 4.5, after 18 time of EM iteration.

Using the arbitrarily selected values, as listed in the third row of Table 4.6, of the parameters, we obtain the estimates of the parameters, which are listed in the fifth row of Table 4.6, after 28 times of EM iteration.

Table 4.6: Comparison among preset, selected and estimated parameters

Parameters	α	β	γ_1	γ_2	q_1	q_2	r	μ_1	μ_2	$RMSD_9$	$RMSD_4$
Exact	0.5	0.8	1.5	1.2	0.32	0.25	0.64	19.2	12	0	0
Selected	0.8	1.2	2	1.5	0.5	0.12	0.45	34	20	5.61	0.38
Change (%)	60	50	33.3	25	56.25	52	29.7	77.1	66.7	17.56	44.25
Estimated	0.83	0.21	1.16	1.44	0.05	0.08	0.06	15.7	15.6	1.71	0.397
Change (%)	66	-73.8	-22.7	20	-0.844	-68	-90.6	-22.3	30	52.65	51.76

4.7.4 RMSD and NRMSD

In Table 4.5, the most data in the fourth row, “change (%)”, are the changes of the guessed values with respect to exact values of the parameters. For the most data in the sixth row, “change (%)”, are for the changes of the estimated values with respect to exact values of the parameters. $RMSD_9$ is about the differences between the guessed or estimated values of nine parameters and the exact values of those values. $RMSD_4$ is about the differences between the guessed or estimated values, of four parameters of interest, α , β , γ_1 and γ_2 , and the exact values of these four parameters of interest. With “ $\hat{\cdot}$ ” to label the guessed or estimated values, the expressions for $RMSD_9$ and $RMSD_4$ are

$$RMSD_9 = \frac{1}{3} \left[(\hat{\alpha} - \alpha)^2 + (\hat{\beta} - \beta)^2 + (\hat{\gamma}_1 - \gamma_1)^2 + (\hat{\gamma}_2 - \gamma_2)^2 + (\hat{q}_1 - q_1)^2 + (\hat{q}_2 - q_2)^2 + (\hat{r} - r)^2 + (\hat{\mu}_1 - \mu_1)^2 + (\hat{\mu}_2 - \mu_2)^2 \right]^{1/2}, \quad (4.6.67)$$

$$RMSD_4 = \frac{1}{2} \left[(\hat{\alpha} - \alpha)^2 + (\hat{\beta} - \beta)^2 + (\hat{\gamma}_1 - \gamma_1)^2 + (\hat{\gamma}_2 - \gamma_2)^2 \right]^{1/2}. \quad (4.6.68)$$

Moreover, the measures for the changes of the values of the nine parameters are $NRMSD_9$ while those for the four parameters are $NRMSD_4$:

$$NRMSD_9 = \frac{1}{3} \left[(\hat{\alpha}/\alpha - 1)^2 + (\hat{\beta}/\beta - 1)^2 + (\hat{\gamma}_1/\gamma_1 - 1)^2 + (\hat{\gamma}_2/\gamma_2 - 1)^2 + (\hat{q}_1/q_1 - 1)^2 + (\hat{q}_2/q_2 - 1)^2 + (\hat{r}/r - 1)^2 + (\hat{\mu}_1/\mu_1 - 1)^2 + (\hat{\mu}_2/\mu_2 - 1)^2 \right]^{1/2}, \quad (4.6.69)$$

$$NRMSD_4 = \frac{1}{2} \left[(\hat{\alpha}/\alpha - 1)^2 + (\hat{\beta}/\beta - 1)^2 + (\hat{\gamma}_1/\gamma_1 - 1)^2 + (\hat{\gamma}_2/\gamma_2 - 1)^2 \right]^{1/2}. \quad (4.6.70)$$

4.7.5 Discussion

We assess the performance of our proposed method based on the comparison among the exact values of parameters, the guessed values and the estimated values. Most values of the guessed parameters are near the values of the preset parameters. This means that the proposed procedure for the guessing of starting parameter values is promising. Among them, the guessed value of q_1 is the least accurate, while the guessed value of the input parameters, γ_1 and γ_2 , are the most accurate.

In Table 4.5, $RMSD_9$ for all the estimated values of the parameters are a bit larger than the $RMSD_9$ for all the guessed values. This is caused mostly by less accurately estimated values of q_1 and q_2 . However, $RMSD_4$ for estimated values is significantly reduced to around half of $RMSD_4$ for guessed value. So at least the estimates of the four parameters of interest are more accurate than the guessed values. Through the comparison of the $NRMSD$ in the fourth row and in the sixth row of Table 4.5, we can confirm that the

approach proposed by us is successful, not only for the four parameters of interest, α , β , γ_1 , γ_2 , but also for all nine parameters.

In Table 4.6, considering most values of the parameters can be between 0 and infinitive, the selected values for the parameters are actually quite near the exact values. However, the RMSDs in Table 4.6 are significantly greater than the RMSDs in Table 4.5. This illustrates that the method we proposed is superior to random initialization in obtaining accurate estimates.

Moreover, more times of EM iteration is need when random initialization is used. This means that random initialization is not only inaccurate but also slower to reach its local optima.

Therefore, we can conclude that the proposed approach on the procedure for starting value guessing and the proposed methods for parameter estimation, is promising. The guessed values are around the preset values. Most of the estimated values, especially the values of the parameters of interest, are much nearer to the preset values than the guessed values are.

Our approach not only demonstrated the accuracy of estimates but also the efficiency and effectiveness of estimation. Although the procedure of the guessing of the starting values of the unknown parameters seems complicated, but the computation is straightforward, no extra time needed for convergence as required by some other initializing methods such as grid search [128]. On the other hand, the EM algorithm is completed within 20 iterations, this verifies Ryden [121]'s conclusion that the EM algorithm is the simplest and quickest method. This also proves the integration of nonlinear mathematical optimization methods, the reducing of the number of unknown parameters

to the nonlinear simultaneous equations, are critical in lower the complexity in computation while saving time consumed for computation.

4.8 Conclusion and Future Research

In this chapter, we propose a state space model to investigate the dynamics of brand equity in Sect. 4.1. This model can separate brand values into two components: BLV (brand label values) and BOV (Brand operation values). BOV is created from a firm's investment and will transfer to BLV through autoregressive process. Both BLV and BOV contribute to brand performance. This is a novel model which can disclose the procedure how brand values are generated.

After the model is proposed, we attempt to identify a dynamic system represented by the state space model. Different with most systems in extant literature, the system to be identified has time-varying exogenous inputs. The Monte Carlo techniques used for the identification of such systems have the issue of residual effect and cost additional computation. This excites us to discover alternative approaches in fulfilling the task of parameter estimation for relevant state space models.

MLE and the EM algorithm are popularly used in parameter estimation for the state space models when the systems have no time-varying exogenous inputs. New approaches must be developed for the application of MLE and the EM algorithm in the identification of a dynamic system. The challenges are mostly on the guessing of the starting values of the parameters, as well as on the optimization in the M step of the EM algorithm, especially when the mathematical optimization is a non-concave nonlinear problem with nonlinear constraints because (1) in practice it can be common that some parameters for state space

model are known and (2) there can be certain nonlinear constraints with respect to the parameters.

To identify such a constrained dynamic system with time-varying exogenous inputs, innovations are needed in the implementation of ELM and the EM algorithm. To guess the starting values of the parameters, we develop a procedure employing different statistic tools and methods, especially the linear regression between the output and the lagged inputs, and the cross covariance between the output and the lagged inputs. During the M step of the EM algorithm, we form another objective function from the expected likelihood and nonlinear constraints. To fulfill the mathematical optimization of the nonlinear objective function with nonlinear constraints, we apply Lagrangian multiplexer method and obtain simultaneous nonlinear equations. Then we choose the trust-region-reflective algorithm to solve these equations. These methods are conventional in mathematical optimization. However, they have never been recommended in the estimation of state space models with time-varying inputs possibly because of the lack of methods to implement the MLE method and the EM algorithms. Therefore, it can be perceived that our innovations to the implementation of EM iteration are mostly on (1) the guessing of the starting value, and (2) the introduction and integration of methods for nonlinear optimization in the M step of the EM algorithm. In order to reduce the computational cost and residue errors, we choose the MLE method, instead of MCMC methods which are popularly chosen for state space models with partly known parameters and exogenous inputs. Hence the application of the MLE methods in state space models with exogenous time-varying inputs and partly known parameters is innovative. In order to apply the MLE methods, we choose the EM algorithm. Because the initial guessing is critical for the accuracy of the estimated parameter values,

the procedure to guess the initial parameter values guarantee the successful implementation of the EM algorithm. In addition, through the introduction and integration of (nonlinear) optimization methods, we finally demonstrate that the application of the MLE methods to obtain accurate estimates of unknown parameters of such a state space model representing brand equity dynamics and to significantly reduce the computational complexity and ineffectiveness is not impossible.

According to the simulation results, in general, the approach proposed, especially the procedure for starting value guessing and the methods for parameter estimation, is promising. The guessed values are around the preset values. Most of the estimated values, especially the values of the parameters of interest, are much nearer to the preset values than the guessed values are. These indicate that the guessing procedure and the estimation methods work fine. However, further investigation on the closeness between the guessed/estimated values, as well as the convergence of the guessing procedure and the estimation methods, need to be performed with theoretical study.

It shall be emphasized that our approach not only demonstrated on the accuracy estimates but also on efficiency and effectiveness. Although the procedure of the guessing of the starting values of the unknown parameters seems complicated, but the computation is straightforward, no extra time needed for convergence as required by most other initializing methods such as grid search [128]. On the other hand, the EM algorithm is completed within 20 iterations, this verifies Ryden [121]'s conclusion that the EM algorithm is the simplest and quickest method. This also proves the integration of nonlinear mathematical optimization methods, the reducing of the number of unknown parameters

to the nonlinear simultaneous equations, are critical in reduces the complexity of computation while saving time consumed for computation.

The approach we are proposing can be employed in future research in identification of dynamic systems represented by state space models, using MLE and the EM algorithm. Further research can be concentrated on certain specific state space models, optimization algorithms, metrics to evaluate the performance of an approach used and applications of certain state space models in practice. For example, models with higher dimensions of exogenous inputs, hidden states and outputs, may reflect the reality of certain applications, such as those in the cases of a firm's marketing investments, brand equity and business success. Moreover, when both objective function and the constraints are more complicated, especially when inequality constraints exist, Karush–Kuhn–Tucker conditions and interior point method, or other methods, must be considered.

4.9 Chapter Summary

As we know, methods to identify state-space-modeled dynamic systems with time-varying exogenous inputs are still limited to algorithms based on Markov chain Monte Carlo which has residue effect and needs additional computation. In this chapter, we intend to identify such dynamic systems using traditional maximum likelihood estimation. This method is often combined with expectation-maximization algorithm, to estimate parameters of state-space-modeled systems without time-varying inputs. It is challenging to fulfill such a task, especially when the parameters of state space models are partly known and the relevant nonlinear objective function constructed is subjected to non-concave nonlinear constraints. In this chapter, we propose an approach, which mainly includes a

procedure developed for initial guessing of values of the parameters, and mathematical optimization methods integrated to obtain the estimates of the parameters in maximize the nonlinear objective function, when a parsimonious state space model for brand equity dynamics is employed. The simulation results demonstrate that our approach, including the guessing procedure developed and the optimization methods integrated, is promising in guessing the starting values and then obtaining the estimates of the parameters.

Chapter V

Conclusions and Future Work

5.1 Conclusions

Generally, our motivation is to construct certain mathematical dynamic models to describe BE dynamics, suitable to the needs of marketing applications. With enough sampling, certain amount of data we intend to obtain insight regarding brand persistence, the effects of advertising and R&D expenses, and the internal relationship among them. Additional attention will be paid to the brand persistence because the long-term effects of the investments on the brands of a firm are of great interest to both executives and managers. The selection of BE measures depends on application need in marketing investments. Since there is much less research using the FBBE categories of approaches than research using CBBE categories of approaches, the FBBE categories of approaches are emphasized for their financial relevance.

However, the extant mathematical models using FBBE to describe BE dynamics cannot reflect the persistence of brand equity. Thus these models are not able to fully represent the dynamics of brand equity. In order to obtain a better representation of the dynamics of brand equity, we propose two types of statistical models: (1) ARX models for the cases where the brand values are known and (2) state space models for cases where the brand values are unknown. Although these models are initially proposed to study the dynamics of FBBE, they can also be used to study the dynamics of CBBE.

Collinearity between exogenous inputs are common in marketing practices. Accordingly, there is collinearity among explanatory variables of a linear regression to estimate an ARX model. Consequently, principal component regression is proposed so we can perform regression on principle components, instead of collinear explanatory variables. These principle components are not collinear because they are orthogonal to each other. Moreover, the number of principle components are significantly less than the number of explanatory variables. This makes it possible to perform linear regression with a limited sample size. Additionally, when there is autocorrelation within the residue series after linear regression, we can further apply the method of generalized differences during the implementation of PCR.

As shown in Chapter II, the application of ARX models helps to perform brand structural analysis and discover insights from the marketing practices of certain firms whose brands are top global brands as claimed by Interbrand. However, to apply ARX models, the brand values must be known. If commercial brand values are not used, the brand values are unobservable by the nature of BE. Consequently, we must propose other models, such as state space models to study BE dynamics.

In existing literature, the identification of state space models through parameters estimation is not well presented yet, especially for the models with nonlinear constraints and exogenous time-varying inputs. The estimation of state space model has been proved to be a task of difficulty and sophistication when the dimension of the states, the output and the inputs are increasing. To make the work less challenging, we not only need break down the tasks, but also need to start from state space models with lower dimensions of inputs, outputs and states. Furthermore, the objective function formatted is usually nonlinear; thus, the guessing of the starting value is critical before the implementation of the algorithms to estimate parameter values.

In this research, we conclude that the ARX models can be estimated using certain proposed approaches. However, due to the collinearity among exogenous inputs, the estimation through ordinary linear regression (OLS) will cause large variance inflation factor (VIF), i.e. the variance of estimate from OLS will become large. Through PCR and generalized differentiation, we can significantly reduce the variance of estimate and the effect of the autocorrelation of the residue errors from PCR, respectively. Consequently, we are able to obtain the parameters of the ARX with acceptable quality, which is measured by the coefficient of determination and p-values.

The application of ARX model in brand structure analysis also helps marketing managers and marketing researchers to reveal more in-depth brand insights than conventional brand outcome measure. We discovered a different brand ranking from Interbrand's brand ranking. We also classified brand management scenarios into four types per their brand intrinsic values and brand performance ratio.

On the other hand, in the estimation of state space models, we conclude that the estimation of state space models with one or two exogenous inputs can be realized through MLE, together with the EM algorithm. In addition, less computational cost is expected for this proposed method than MCMC-based methods. However, in order to implement the proposed MLE method successfully, it is critical to carefully design the procedure to guess the initial parameter values to be estimated and the approaches to fulfill the M step of the EM algorithm. Our simulation results indicate that the initial guessing procedure developed and mathematical optimization methods proposed are promising in estimating state space models.

5.1.1 Contributions

This research is an important attempt to identify dynamic systems represented by ARX models or state space models. In Chapters II, III and IV, the contributions of the research in each chapter have been introduced. The highlights of the contributions are stated as below.

In Chapter II, through the successful application of ARX model and the use of Compustat and Interbrand data, an innovative generic brand value structure analysis is performed. Consequently, brand outcome can be used for brand diagnostic purposes and to qualify the long-term brand value and assess firm brand operation. Theoretically, we reached the target to separate the long-term effects and short-term effects. Besides, we recommend intrinsic brand value and performance ratio in brand assessment which are important to obtain a profile on the status of a firm's brand. Managerially, through their understanding of the structure of brand value, firms' decision makers are able to recognize

the difference among the effects of investments on advertising and R&D, in order to make smart decision to optimize the returns from marketing investments.

In Chapter III, we find that there is a relationship between the variance of the initial state and the variance of the process noise so we don't need to estimate both. This finding is used in Chapter IV with some variation, which is caused by the difference of the state space models used in these two chapters. Even more, the proposed the WLS method for the initial guessing of parameter values is innovative. The method works for a state space model which has only one dimension of input, state and output, respectively.

In Chapter IV, to reduce computational complexity and cost, MLE, together with the EM algorithm, is proposed for the estimation of a two-input state space model with nonlinear constraints and time-varying exogenous inputs. Since the elements of the parametric matrices are partly known, we are not able to use the traditional way of treating whole matrices as unknown parametric matrices. Moreover, the time-varying inputs and nonlinear constraints greatly increased the complexity of the mathematical optimization problem derived from MLE. Accordingly, not only the objective function itself is nonlinear but also the constraints to the objective function are nonlinear. Such an optimization problem to be solved by the EM algorithm is rarely seen in extant literature. In order to implement the EM algorithm, an innovative procedure for the initial guessing of the parameter values is designed and an integrated optimization approach for parameter estimation is developed. Through simulation, the procedure and the approach are indicated to be effective in obtaining the parameter values and thus estimating such a state space model.

5.2 Future Work

In Chapter II, during the application of the ARX model, we assumed that the brand intrinsic value remains stable over the observation period. However, in practice, brand intrinsic value may change over time (e.g., in response to market changes and brand investments). Further research can model a changing brand intrinsic value, in response to market inputs and brand performance, using advanced modeling techniques. Such a model can also be a state space model, where the brand intrinsic value can be treated as an observable state which is evolving over time.

The further research of Chapter III is performed in Chapter IV. The approaches we proposed in these two chapters can be employed in future research in the identification of dynamic systems represented by state space models, using MLE and the EM algorithm. Further research can be concentrated on (1) certain specific state space models for certain industrial applications (eg. models with higher dimensions of exogenous inputs, hidden states and outputs than the model used in Chapter IV, may represent the reality of certain applications, such as those in the case of a firm's marketing investments, brand equity and business success); (2) optimization algorithms to include inequality constraints (eg., when both objective function and the constraints are more complicated, especially when inequality constraints exist, Karush–Kuhn–Tucker conditions and interior point method, or other methods, will be considered); (3) metrics on efficiency or convergence to carefully evaluate the performance of a proposed approach used, and (4) applications of certain state space models in practice with sufficient financing, accounting and/or customer-surveying data provided by a certain company.

Other than brand equity, customer satisfaction [137] and customer equity [138] [139] are the topics closely related to CBBE. Like brand equity, both are also latent variables in marketing research. Analytical tools such as latent growth structural equation modelling (SEM) are popular to handle problems regarding customer satisfaction and loyalty. However, SEM is stronger in theory testing than decision making [140] - [142], so they are not a good alternative to state space models. In the future, the combination of SEM and state space model can be considered.

Bibliography

- [1] AMA, (2016) *Dictionary* [online], available:
<https://www.ama.org/resources/pages/dictionary.aspx?dLetter=B>.
- [2] P. H. Farquhar, "Managing brand equity," *Marketing Research*, vol. 1, pp. 24-33,
September 1989.
- [3] K. L. Keller, *Strategic Brand Management*, Prentice Hall, New Jersey, 1998.
- [4] J. C. Crimmins, "Better measurement and management of brand value," *Journal of
Advertising Research*, vol. 40, no. 6, pp. 136-144, 2000.
- [5] D. A. Aaker, *Building Strong Brands*, The Free Press, New York, 1996.
- [6] M. Neumeier, *The Brand Gap: How to Bridge the Distance Between Business
Strategy and Design*, Berkeley, CA: New Riders Publishing, 2006.
- [7] K. L. Keller, "Building and managing corporate brand equity," in *The Expressive
Organization*, eds. Majken Schultz, Mary Jo Hatch, Oxford University Press, 2000.
- [8] G. Kleisterlee, "Innovation as driver of sustainable growth," *Speech at China Central
Party School*, 2007.
- [9] D. A. Aaker, *Managing Brand Equity*. The Free Press, New York, 1991.

- [10] K. L. Keller, "Conceptualizing measuring, and managing customer-based brand equity," *Journal of Marketing*, vol. 57, no.1, pp. 1-22, 1993.
- [11] C. S. Park and V. Srinivasan, "A survey-based method for measuring and understanding brand equity and its extendibility," *Journal of Marketing Research*, vol. 39, no. 4, pp. 421-439, 1994.
- [12] V. Srinivasan, C. S. Park, and D. R. Chang, "An approach to the measurement, analysis, and prediction of brand equity and its sources," *Management Science*, vol. 51, no. 9, pp. 1433-1448, 2005.
- [13] W. Kamakura and G. Russell, "Measuring brand value with scan data," *International Journal for Research in Marketing*, vol.10, no. 1, pp. 9-22, 1993.
- [14] C. J. Simon and M. W. Sullivan, "The measurement and determinants of brand equity: A financial approach," *Marketing Science*, vol. 12, no.1, pp. 28-52, 1993.
- [15] V. Mahajan, S. Sharma, and R. D. Buzzell. "Assessing the impact of competitive entry on market expansion and incumbent sales." *Journal of Marketing*, vol. 57, pp. 39-52, 1993.
- [16] D. A. Aaker, "Measuring brand equity across products and markets," *California Management Review*, vol. 38, pp. 102-120, Spring 1996.
- [17] K. L. Ailawadi, D. R. Lehmann, and S. A. Neslin, "Revenue premium as an outcome measure of brand equity," *Journal of Marketing*, vol. 67, pp. 1-17, 2003.
- [18] S. Chu and H. T. Keh, "Brand value creation: Analysis of the Interbrand-Business Week brand value rankings," *Marketing Letters*, vol. 17, no. 4, pp. 323-331, 2006.

- [19] D. M. Hanssens and M. G. Dekimpe, "Modeling the financial-performance effects of marketing," in B. Wierenga (Ed.), *Handbook of Marketing Decision Models*, pp. 501-523, 2008.
- [20] D. A. Aaker, (2011), *When will a damaged brand come back?* [online], available: <http://www.prophet.com/blog/aakeronbrands/45-when-will-a-damaged-brand-come-back>.
- [21] A. S. Dick and K., Basu, "Customer loyalty: Toward an integrated conceptual framework," *Journal of the Academy of Marketing Science*, vol. 22, no. 2, pp. 99-113, 1994.
- [22] V. Pare and J. Dawes, "The persistence of excess brand loyalty over multiple years," *Marketing Letters*, vol. 23, no. 1, pp. 163-175, 2012.
- [23] M. P. Conchar, M. R. Crask, and G. M., Zinkhan, "Market valuation models of the effect of advertising and promotional spending: a review and meta-analysis," *Journal of the Academy of Marketing Science*, vol. 33, no. 4, pp. 445-460, 2005.
- [24] Y.-M. Chen, "The persistence of brand value at country, industry, and firm levels," Special Issue: Brand Equity, Branding, and Marketing Communications in Emerging Markets. *Journal of Global Marketing*, vol. 23, no.3, pp. 253-269, 2010.
- [25] K. Jedidi, C. F. Mela, and S. Gupta, "Managing advertising and promotion for long-run profitability," *Marketing Science*, vol. 18, no. 1, pp. 1-22, 1999.
- [26] F. Wang, X.-P. Zhang, and M. Ouyang, "Does advertising create sustained firm value? The capitalization of brand intangible," *Journal of the Academy of Marketing Science*, vol. 37, no.2, pp. 130-143, 2009.

- [27] I., Buila, L. de Chernatony, and E. Martíneza, “Examining the role of advertising and sales promotions in brand equity creation,” *Journal of Business Research*, vol. 66, no.1, pp.115-122, 2013.
- [28] K. Pauwels, D. Hanssens, and S. Siddharth, “The long-term effects of price promotions on category incidence, brand choice, and purchase quantity,” *Journal of Marketing Research*, vol. 39, pp. 421-439, 2002.
- [29] A. D’Astous and I. Jacob, “Understanding consumer relations to premium-based promotional offers,” *European Journal of Marketing*, vol. 36, vol. 11/12, pp. 1277-1287, 2002.
- [30] G. S. Low and J. J. Mohr, “Advertising vs sales promotion: a brand management perspective,” *Journal of Product & Brand Management*, vol. 9, no. 6, pp. 389-414, 2000.
- [31] N. Sedaghat, M. Sedaghat, and A. K. Moakher, “The impact of promotional mix elements on brand equity,” *American Journal of Scientific Research*. vol. 43, pp. 5-15, 2012.
- [32] R. Kline, *Principles and Practice of Structural Equation Modeling*, 3rd Ed., Guilford, 2011.
- [33] J. C. Westland, *Structural Equation Modeling: From Paths to Networks*, New York: Springer, 2015
- [34] NA, “World-wide R&D: Widening recognition of R&D importance,” *Strategic Direction*, vol. 22, no.3, pp.30-32, 2006.

- [35] W. R. Dillon, T. J. Madden, A. Kirmani, and S. Mukherjee, "Understanding what's in a brand rating: A model for assessing brand and attribute effects and their relationship to brand equity," *Journal of Marketing Research*, vol. 38, no. 4, pp. 415-429, November 2001.
- [36] S. Sriram, S. Balachander, and M. U. Kalwani, "Monitoring the dynamics of brand equity using store-level data," *Journal of Marketing*, vol. 71, pp. 61-78, April 2007.
- [37] V. Shankar, P. Azar, and M. Fuller, "BRAN*EQT: A multicategory brand equity model and its application at Allstate," *Marketing Science*, vol. 27, no.4, pp. 567-84, 2008.
- [38] A. Aribarg and N. Arora, "Interbrand variant overlap: Impact on brand preference and portfolio profit," *Marketing Science*, vol. 27, nol. 3, pp. 474-491, 2008.
- [39] S. Voleti and P. Ghosh, "A non-parametric model of residual brand equity in hierarchical branding structures with application to US beer data," *Serie A, Statistics in Society*, vol. 177, no. 1, pp. 135–152, January 2014.
- [40] J. Tobin (1969), "A general equilibrium approach to monetary theory," *Journal of Money, Credit and Banking*, vol. 1, no. 1, pp. 15–29, 1969.
- [41] L. Ljung, *System Identification: Theory for the User*, 2nd Ed., PTR Prentice Hall, Upper Saddle River, N.J., 1999.
- [42] J. H. Roberts, U. Kayande, and S. Stremersch, "From academic research to marketing practice: Exploring the marketing science value chain," *International Journal of Research in Marketing*, vol. 31, no. 2, pp. 127-140, 2014.

- [43] K. L. Keller and D. R. Lehmann, "Brands and branding: Research findings and future priorities," *Marketing Science*, vol. 25, no. 6, pp. 740–759, 2006.
- [44] C. Eckert, J. L. Jordan, and T. Islam, "Seeing the forest despite the trees: Brand effects on choice uncertainty," *International Journal of Research in Marketing*, vol. 29, no. 3, pp. 256-264, 2012.
- [45] H. Guyon and J.-F. Petiot, "New conjoint approaches to scaling brand equity and optimising share of preference prediction," *International Journal of Market Research*, vol. 57, no. 5, pp. 701-726, 2015.
- [46] Interbrand, (2016), *Best global brands 2016* [online], available : <http://www.interbrand.com/en/best-global-brands/2013/Best-Global-Brands-2016.aspx>, 2016.
- [47] V. Srinivasan, C. S. Park, and D. R. Chang, "An approach to the measurement, analysis, and prediction of brand equity and its sources," *Management Science*, vol. 51, no.9, pp. 1433-48, 2005.
- [48] N. S. Davcik and P. Sharma, "Impact of product differentiation, marketing investments and brand equity on pricing strategies: A brand level investigation," *European Journal of Marketing*, vol. 49, no. 5/6, pp. 760-781, 2015.
- [49] R. T. Rust, T. Ambler, G. S. Carpenter, V. Kumar, and R. K. Srivastava, "Measuring marketing productivity: Current knowledge and future directions," *Journal of Marketing*, vol. 68, pp. 76-89, 2004.

- [50] J. L. Zaichkowsky, M. Parlee, and J. Hill, "Managing industrial brand equity: Developing tangible benefits for intangible assets," *Industrial Marketing Management*, vol. 39, no. 5, pp. 776-783, 2010.
- [51] R. N. Sinclair and K. L. Keller, "A case for brands as assets: Acquired and internally developed," *Journal of Brand Management*, vol. 21, no. 4, pp. 286-302, 2014.
- [52] R. Amit and P. J. H. Schoemaker, "Strategic assets and organizational rent," *Strategic Management Journal*, vol. 14, no. 1, pp. 33-46, 1993.
- [53] J. K. Johansson, C. V. Dimofte, and S. K. Mazvancheryl, "The performance of global brands in the 2008 financial crisis: A test of two brand value measures," *International Journal of Research in Marketing*, vol. 29, no.3, pp. 235-245, 2012.
- [54] F. J. Hsu, T. Y. Wang, and M. Y. Chen, "The Impact of brand value on financial performance," *Advances in Management and Applied Economics*, vol. 3, no.6, pp. 129-141, 2013.
- [55] M., Dutordoir, F. H. M. Verbeeten, and D. De Beijer, "Stock price reactions to brand value announcements: Magnitude and moderators," *International Journal of Research in Marketing*, vol. 32, no.1, pp. 34-47, 2015.
- [56] J. Ratnatunga, and M. T. Ewing, "An ex-ante approach to brand capability valuation," *Journal of Business Research*, vol. 62, no.3, pp.323-331, 2009.
- [57] Y.-C. J. Wu, "Renaming effect of brand value: state-owned enterprises," *Management Decision*, vol. 47, no.10, pp. 1555-1581, 2009.

- [58] E. Papista and S. Dimitriadis, "Exploring consumer-brand relationship quality and identification," *Qualitative Market Research: An International Journal*, vol. 15, no. 1, pp. 33-56, 2012.
- [59] W. Fritz, B. Lorenz, and M. Kempe, "An extended search for generic consumer–brand relationships," *Psychology & Marketing*, vol. 31, no.11, pp. 976-991, 2014.
- [60] S. Hudson, L. Huang, M. S. Roth, and T. J. Madden, "The influence of social media interactions on consumer–brand relationships: A three-country study of brand perceptions and marketing behaviors," *International Journal of Research in Marketing*, vol. 33, no. 1, pp. 27-41, 2016.
- [61] Y. Odin, N. Odin, and P. Valette-Florence, "Conceptual and operational aspects of brand loyalty: An empirical investigation," *Journal of Business Research*, vol. 53, no. 2, pp. 75–84, 2001.
- [62] L.-W. Wu, "Satisfaction, inertia, and customer loyalty in the varying levels of the zone of tolerance and alternative attractiveness," *Journal of Services Marketing*, vol. 25, no. 5, 310-322, 2011.
- [63] F. Jørgensen, T. A. Mathisen, and P. Hassa, "Brand loyalty among Norwegian car owners," *Journal of Retailing and Consumer Services*, vol. 31, pp. 256-264, 2016.
- [64] J. Kim, J. D. Morris, and J. Swait, "Antecedents of true brand loyalty," *Journal of Advertising*, vol. 37, no. 2, pp. 99-117, 2008.
- [65] J. Lee and H. Lee, "Does satisfaction affect brand loyalty?" *Academy of Marketing Studies Journal*, vol.17, no. 2, pp. 488-500, 2013.
- [66] P. Kotler, *Marketing Management*, 7th Ed., Prentice-Hall, 1991.

- [67] M. R. Solomon, G. Bamossy, and S. Askegaard, *Consumer Behaviour: A European Perspective*, 2nd Ed., Harlow, Essex: Pearson Education, 2002.
- [68] A. K. Paswan, N. Spears, and G. Ganesh, "The effects of obtaining one's preferred service brand on consumer satisfaction and brand loyalty," *Journal of Services Marketing*, vol. 21, no. 2, pp. 75-87, 2007.
- [69] H. H. Chang and L. H. Wu, "An examination of negative e-WOM adoption: Brand commitment as a moderator," *Decision Support Systems*, vol. 59, pp. 206-218, 2014.
- [70] K. Hung, S. Y. Li, and D. K. Tse, "Interpersonal trust and platform credibility in a Chinese multibrand online community: Effects on brand variety seeking and the time spent," *Journal of Advertising*, vol. 40, no. 3, pp. 99-112, 2011.
- [71] J.-P. Dubé, G. J. Hitsch, P. E. Rossi, "State dependence and alternative explanations for consumer inertia," *The RAND Journal of Economics*, vol. 41, no. 3, pp. 417-445, 2010.
- [72] J. T. Prince, "Relating inertia and experience in technology markets: An analysis of households' personal computer choices," *Applied Economics*, vol. 43, no. 29, pp. 4501-4514, 2011.
- [73] J. Fan, P. Venkat, R. Gulati, and V. Kumar, "Marketing-mix recommendations to manage value growth at P&G Asia-Pacific," *Marketing Science*, vol. 28, no. 4, pp. 645-655, 2009.
- [74] N. Brooks and L. Simkin, "Judging marketing mix effectiveness," *Marketing Intelligence & Planning*, vol. 30, no. 5, pp. 494 - 514, 2012.

- [75] C. P. Kirk, I. Ray, and B. Wilson, "The impact of brand value on firm valuation: The moderating influence of firm type," *Journal of Brand Management*, vol. 20, no. 6, pp. 488-500, June, 2013.
- [76] B. J. Bronnenberg, S. K. Dhar, and J.-P. H. Dubé, "Brand history, geography, and the persistence of brand shares," *Journal of Political Economy*, vol. 117, no. 1, pp. 87-115, 2009
- [77] L. L. Eng and H. T. Keh, "The effects of advertising and brand value on future operating and market performance," *Journal of Advertising*, vol. 36, no. 4, pp. 91-100, 2007.
- [78] R. A. Peterson and J. Jeong, "Exploring the impact of advertising and R&D expenditures on corporate brand value and firm-level financial performance," *Journal of the Academy of Marketing Science*, vol. 38, no. 6, pp. 677–690, 2010.
- [79] A. Torres and J. A. Tribó, "Customer satisfaction and brand equity," *Journal of Business Research*, vol. 64, no. 10, pp. 1089–1096, 2011.
- [80] I. T. Jolliffe, "Principal component analysis," *Series: Springer Series in Statistics*, 2nd Ed., New York: Springer, 2002.
- [81] E. R., Mansfield, "PCR: Principal component regression analysis," *Journal of Marketing Research*, vol. 15, pp. 471-72, 1978.
- [82] G. T. Knofczynski and D. Mundfrom, "Sample size when using multiple linear regression for prediction," *Educational and Psychological Measurement*, vol. 68, pp. 431–442, June 2008.

- [83] S. Srinivasan, K., Pauwels, J. Silva-Risso, and D. M. Hanssens, “Product innovations, advertising, and stock returns,” *Journal of Marketing*, vol. 73, pp. 24-43, January 2009.
- [84] W. H. Greene, *Econometric Analysis*, 7th Ed., Pearson, New Jersey: Prentice Hall, 2012.
- [85] Consumer Reports, (2010) *Survey: Toyota crisis taking its tolls on brand loyalty* [online], available: <http://www.consumerreports.org/cro/news/2010/02/survey-toyota-crisis-taking-its-toll-on-brand-loyalty/index.htm>.
- [86] M. Kariya, (2004), *Working on two wheels* [online], available: <http://www.policeone.com/police-products/vehicle-equipment/articles/79717-Working-on-two-wheels/> accessed on Oct. 18, 2013, 2004.
- [87] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey, 1996.
- [88] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1-38, 1977.
- [89] Z. Ghahramani and G. E. Hinton, “Parameter estimation for linear dynamic systems,” *Technical report CRC-TR-96-2*, Department of Computer Science, University of Toronto, 1996.
- [90] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples*, Springer New York, 2011.

- [91] E. E. Holmes, "Derivation of an EM algorithm for constrained and unconstrained multivariate autoregressive state-space (MARSS) models," *Technical Report*, Northwest Fisheries Center, NOAA Fisheries, 2013.
- [92] L. S.-Y. Wu, J. S. Pai, and J. R. M. Hosking, "An algorithm for estimating parameters of state-space models," *Statistics & Probability Letters*, vol. 28, pp. 99-106, 1996.
- [93] C. F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, Vol. 11, No. 1, pp. 95-103.
- [94] A. F. Zuur, R. J. Fryer, I. T. Jolliffe, R. Dekker, and J. J. Beukema, "Estimating common trends in multivariate time series using dynamic factor analysis," *Environmetrics*, vol. 14, no. 7, pp. 665-685, 2003.
- [95] A. Harvey and S. J. Koopman, "Unobserved components models in economics and finance," *IEEE Control Systems Magazine*, vol. 29, no. 6, pp. 71-81, 2009.
- [96] P. Tendeo, P. Ailliot, and E. Autret, "Linear Guassian state-space model with irregular sampling: application to sea surface temperature," *Stochastic Environmental Research and Risk Assessment*, vol. 25, pp. 793-804, 2011.
- [97] H. J. Van Heerde, C. Mela, and P. Manchanda, "The dynamic effect of innovation on market structure," *Journal of Marketing Research*, vol. 41, no. 2, pp. 166-183, 2004.
- [98] M. B. Ataman, C. F. Mela, and H. J. Van Heerde, "Building brands," *Marketing Science*, vol. 27, no. 6, pp. 1036-1054, 2007.

- [99] N. Bruce and Z. Foutz, "Dynamic advertising and word-of-mouth effectiveness in sequential distribution of short lifecycle products," *Marketing Dynamics Conference*, 2007.
- [100] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82. no. 1, pp. 35–45, 1960.
- [101] M. Kotani and U. Sumita, "Structural analysis of "national brand vs. store brand" with stochastic demands," *International Journal of Business and Information*, vol. 8, no. 1, pp. 1-33, 2013.
- [102] P. Ostergaard, J. Hermansen, and J. Fitchett, "Structures of brand and anti-brand meaning: a semiotic square analysis of reflexive consumption," *Journal of Brand Management*, vol. 22, no. 1, pp. 60-77, 2015.
- [103] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time series Analysis*, vol. 3, no. 4, pp. 253-264, 1982.
- [104] L. Deng and X. Shen, "Maximum likelihood in statistical estimation of dynamic systems: decomposition algorithm and simulation results," *Signal Processing*, vol. 57, pp. 65-79, 1997.
- [105] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667-1682, 2005.
- [106] T. B. Schon, A. Wills, and B. Ninness, "Maximum likelihood nonlinear system estimation," *Proceedings 14th IFAC Symposium on System Identification (SYSID)*, 2009.

- [107] J. Kokkala, A. Solin, and S. Sarkka, "Expectation maximization based parameter estimation by sigma-point and particle smoothing," *17th International Conference on Information Fusion (FUSION)*, 2014.
- [108] G. A. Einicke, G. Falco, M. T. Dunn, and D. C. Reid, "Iterative smoother-based variance estimation," *IEEE Signal Processing Letters*, vol. 19, no. 5, pp. 275-278, 2012.
- [109] C. Andrieu, N. De Freitas, and A. Doucet, "An introduction to MCMC for machine learning," *Machine Learning*, vol. 50, pp. 5-43, 2003.
- [110] J. C. Spall, "Estimation via Markov chain Monte Carlo," *IEEE Control Systems Magazine*, vol. 23, no. 2, pp. 34-45, April 2003.
- [111] J. Hobert, "The data augmentation algorithm," *The Handbook of Markov Chain Monte Carlo*, CRC Press, 2008.
- [112] O. Papaspiliopoulos, G. O. Roberts, and M. Sköld, "A general framework for the parameterization of hierarchical models," *Statistical Science*, vol. 22, pp. 59-73, 2007.
- [113] A. Lee and N. Whiteley, "Variance estimation and allocation in the particle filter," *Working paper* [online], available: <https://arxiv.org/pdf/1509.00394v1.pdf>, 2016.
- [114] O. Cappe, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899-924, 2007.

- [115] P. Del Moral and F.-K. Formulae, “Genealogical and interacting particle systems with applications,” *Series: Probability & Applications*, Springer Verlag; New York: Springer-Verlag, 2004.
- [116] T. Flury and N. Shephard, “Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models,” *Econometric Theory*, vol. 27, pp. 933–956, 2011.
- [117] J. Olsson, O. Cappe, R. Douc and E. Moulines, “Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models,” *Bernoulli* vol. 14, no. 1, pp. 155-179, 2008,
- [118] N. Polson, J. Stroud and P. Mueller, “Practical filtering with sequential parameter learning,” *Journal of Royal Statistical Society, Series B (Statistical Methodology)*, vol. 70, pp. 413-128, 2008.
- [119] A. Belloni, V. Chernozhukov, “On the computational complexity of MCMC-based estimators in large sample,” *The Annals of Statistics*, vol. 37, no. 4, pp. 2011-2055, 2009.
- [120] P. S. H. Leeflang, T. H. A. Bijmolt, J. van Doorn, D. M. Hanssens, H. J. van Heerde, P. C. Verhoef, and J. E. Wieringa, “Creating lift versus building the base: Current trends in marketing dynamics,” *International Journal of Research in Marketing*, vol. 26, no. 1, pp. 13–20, 2009.
- [121] T. Ryden, “EM versus Markov chain Monte Carlo for estimation of hidden Markov models: a computational perspective,” *Bayesian Analysis*, vol. 3, no. 4, pp. 659-688, 2008.

- [122] G. J. McLachlan and T. Krishnam, *The EM algorithm and Extensions*, Wiley, New York, 1997.
- [123] M. Kuroda and M. Sakakihara, "Accelerating the convergence of the EM algorithm using the vector ε algorithm," *Computational Statistics and Data Analysis*, vol. 51, no. 3, pp. 1549-1561, 2006.
- [124] F.-X. Jollois and M. Nadif, "Speed-up for the expectation-maximization algorithm for clustering categorical data." *Journal of Global Optimization*, vol. 37, no. 4, pp. 513- 525, 2007.
- [125] A. Fanelli, M. Flammini, and L. Moscardelli, "The speed of convergence in congestion games under best-response dynamics," *ACM Transactions on Algorithms (TALG)*, vol. 8, no. 3, pp. 1-5, 2012.
- [126] T. Flury and N. Shephard, "Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models," *Econometric Theory*, Cambridge University Press, vol. 27, no. 5, pp. 933-956, October 2011.
- [127] W. Seidel, K. Mosler, and M. Alker, "A cautionary note on likelihood ratio tests in mixture models," *Annals of the Institute of Statistical Mathematics*, vol. 52, pp. 481–487, 2000.
- [128] D. Karlis and E. Xekalaki, "Choosing initial values for the EM algorithm for finite mixtures," *Computational Statistics and Data Analysis*, vol. 41, pp. 577-590, 2003.
- [129] S. Frauke, "Finding starting-value for maximum likelihood estimation of vector STAR Models," *Discussion Paper No. 13-076*, Center for European Economic Research, 2015.

- [130] J. Bergstra, and Y. Bengio, “Random search for hyper-parameter optimization,” *Journal of Machine Learning Research*, vol. 13, pp. 281-305, 2012.
- [131] K. O. Bowman, L. R. Shenton, “Estimator: method of moments,” *Encyclopedia of statistical sciences*, Wiley, pp. 2092-2098, 1998.
- [132] W. D. Furman and B. G. Lindsay, “Measuring the relative effectiveness of moment estimators as starting values in maximizing likelihoods,” *Computational Statistics and Data Analysis*, vol.17, pp. 493–508, 1994.
- [133] E. Bibbona, G. Panfilo, and P. Tavella, “The Ornstein-Uhlenbeck process as a model of a low pass filtered white noise,” *Metrologia*, vol. 45, no. 6, pp. 117-126, 2008.
- [134] C. Biernackia, G. Celeux, and G. Govaert, “Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models,” *Computational Statistics & Data Analysis*, vol. 41, pp. 561-575, 2003.
- [135] Y.-X. Yuan, “Recent advances in trust region algorithms,” *Mathematical Programming*, vol. 151, no. 1, pp. 249-281, 2015.
- [136] M. A. Branch, T. F. Coleman, and Y. Li, “A Subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems,” *SIAM Journal on Scientific Computing*, vol. 21, no. 1, pp. 1-23, 1999.
- [137] C. Fornell, F. V. Morgeson III, and Hult, G. T. M., “An abnormally abnormal intangible: stock returns on customer satisfaction,” *Journal of Marketing*, vol. 80, no. 5, pp. 122-125, 2016.
- [138] R. T. Rust, J. Kim, Y. Dong, T. J. Kim, and S. Lee, “Drivers of customer equity,” *Handbook of Research on Customer Equity in Marketing*, vol. 17, 2015.

- [139] R. T. Rust and M. H. Huang, "The service revolution and the transformation of marketing science," *Marketing Science*, vol. 33, no. 2, pp. 206-221, 2014.
- [140] J.-B. E. M. Steenkamp and H. Baumgartner, "On the use of structural equation models for marketing modeling," *International Journal of Research in Marketing*, vol. 17, pp. 195-202, 2000
- [141] R. Bagozzi, Y. Yi, "Specification, evaluation, and interpretation of structural equation models," *Journal of the Academy of Marketing Science*, vol. 40, no. 1, pp. 8-34, 2012.
- [142] J. C. Westland, *Structural Equation Modeling: From Paths to Networks*. New York: Springer, 2015