

MCMC METHODS FOR WAVELET REPRESENTATIONS IN SINGLE INDEX
MODELS

A Dissertation

by

CHUN GUN PARK

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2003

Major Subject: Statistics

MCMC METHODS FOR WAVELET REPRESENTATIONS IN SINGLE INDEX
MODELS

A Dissertation

by

CHUN GUN PARK

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Jeffrey D. Hart
(Co-Chair of Committee)

Marina Vannucci
(Co-Chair of Committee)

Bani Mallick
(Member)

Dante DeBlassie
(Member)

James A. Calvin
(Head of Department)

August 2003

Major Subject: Statistics

ABSTRACT

MCMC Methods for Wavelet Representations in Single Index Models. (August 2003)

Chun Gun Park , B.S., Kyonggi University, Korea;

M.A., ChungAng University, Korea

Co-Chairs of Advisory Committee: Dr. Jeffrey D. Hart
Dr. Marina Vannucci

Single index models are a special type of nonlinear regression model that are partially linear and play an important role in fields that employ multidimensional regression models. A wavelet series is thought of as a good approximation to any function in the L^2 space. There are two ways to represent the function: one in which all wavelet coefficients are used in the series, and another that allows for shrinkage rules. We propose posterior inference for the two wavelet representations of the function.

To implement posterior inference, we define a hierarchial (mixture) prior model on the scaling(wavelet) coefficients. Since from the two representations a non-zero coefficient has information about the features of the function at a certain scale and location, a prior model for the coefficient should depend on its resolution level. In wavelet shrinkage rules we use "pseudo-priors" for a zero coefficient.

In single index models a direction β affects estimates of the function. Transforming β to a spherical polar coordinate θ is a convenient way of imposing the constraint $\|\beta\| = 1$. The posterior distribution of the direction is unknown and we employ a Metropolis algorithm and an independence sampler, which require a proposal distribution. A normal distribution is proposed as the proposal distribution for the direction. We introduce ways to choose its mode (mean) using the independence sampler.

For Monte Carlo simulations and real data we compare performances of the Metropo-

lis algorithm and independence samplers for the direction and two functions: the cosine function is represented only by the smooth part in the wavelet series and the Doppler function is represented by both smooth and detail parts of the series.

To my family

ACKNOWLEDGMENTS

First and foremost, I wish to thank to the Lord Jesus Christ for completion of my dissertation. He always guides me to the right way, and always opens my eyes to "my unrealized potential".

I would like to thank Dr. Jeffrey D. Hart and Dr. Marina Vannucci, my co-advisors, for their guidance and encouragement. I will never forget their countless hours of advice and direction. Dr. Jeffrey D. Hart expressed his interest in what I researched and guided me to a better perspective on my own work. Dr. Marina Vannucci instructed me with her knowledge of wavelet analysis in statistics and supported me financially.

I am also thankful to both of my committee members, Dr. Bani Mallick and Dr. Dante DeBlassie, for their suggestions.

I am grateful to my parents and my sisters for their support, patient, and *love*. Without them, my dissertation would never have come into existence.

Finally, I wish to thank my wife, Hanmee, and my sons, John and Nathanael, for constant understanding and encouraging me in my Ph.D. journey.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGMENTS	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER	
I INTRODUCTION	1
1.1 Single index models	1
1.2 Bayesian wavelet methods for single index models	2
1.3 Outline of the dissertation	3
II BAYESIAN METHODS FOR WAVELETS	4
2.1 Wavelet series	4
2.2 Bayesian approaches to wavelet series	5
III BAYESIAN METHODS FOR WAVELET REPRESENTATIONS OF SINGLE INDEX MODELS	7
3.1 Wavelet representations of single index models	7
3.2 Likelihood function and prior distributions	7
3.3 Posterior inference and MCMC schemes	13
3.4 An independence sampler for a normal proposal distribution . .	19
3.5 Computational issues	21
IV SIMULATION STUDIES AND AN APPLICATION	25
4.1 Simulation study for fixed resolution	25
4.2 Simulation study for wavelet shrinkage	51
4.3 A real data example	67

CHAPTER	Page
V CONCLUSIONS	73
5.1 Summary	73
5.2 Future work	73
REFERENCES	74
APPENDIX A	77
APPENDIX B	79
APPENDIX C	83
VITA	88

LIST OF TABLES

TABLE	Page
1 Simulation results for $y_i = \cos(z_i) + \varepsilon_i$ where $z_i = X^i T(\theta)$. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined based on an acceptance rate (see Subsection 3.5.2).	27
2 Simulation results for $y_i = 2\sqrt{z_i(1-z_i)} \sin\left(\frac{2.1\pi}{z_i+0.05}\right) + \varepsilon_i$ with the maximum support and with no shrinkage where $z_i = X^i T(\theta)$. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined based on an acceptance rate (see Subsection 3.5.2).	34
3 Simulation results for $y_i = 2\sqrt{z_i(1-z_i)} \sin\left(\frac{2.1\pi}{z_i+0.05}\right) + \varepsilon_i$ with the adjusted support and with no shrinkage where $z_i = X^i T(\theta)$. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined based on an acceptance rate (see Subsection 3.5.2).	43
4 Simulation results for $y_i = 2\sqrt{z_i(1-z_i)} \sin\left(\frac{2.1\pi}{z_i+0.05}\right) + \varepsilon_i$ with the maximum support and with shrinkage where $z_i = X^i T(\theta)$. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined based on an acceptance rate (see Subsection 3.5.2).	51
5 Simulation results for $y_i = 2\sqrt{z_i(1-z_i)} \sin\left(\frac{2.1\pi}{z_i+0.05}\right) + \varepsilon_i$ with the adjusted support and with shrinkage where $z_i = X^i T(\theta)$. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined based on an acceptance rate (see Subsection 3.5.2).	59
6 Results on air pollution data	67

LIST OF FIGURES

FIGURE	Page
1	Circular polar coordinates. 9
2	Spherical polar coordinates. 9
3	A single index model without shrinkage as a graphical model. 17
4	A single index model with shrinkage as a graphical model. 19
5	Use three points to find a mode (mean). The dashed line is proportional to the target distribution of the direction, given all other parameters. 22
6	Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command <i>fminsearch</i> (dash-dot line), together with the true curve(solid line) on the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.02^2)$ 28
7	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.02^2)$. The box has lines at the lower quartile, median, and upper quartile values. 28
8	Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.5^2)$ 29
9	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.5^2)$. The box has lines at the lower quartile, median, and upper quartile values. 29
10	Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 1^2)$ 30

- 11 Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 1^2)$. The box has lines at the lower quartile, median, and upper quartile values. 30
- 12 Histograms for the posterior distribution of $c_{0,k}$ for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.02^2)$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000. 31
- 13 Histograms for the posterior distribution of $c_{0,k}$ for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.5^2)$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000. 32
- 14 Histograms for the posterior distribution of $c_{0,k}$ for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 1^2)$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000. 33
- 15 Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence based on two lines (dashed line), and the independence based on the Matlab command *fminsearch* (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 0.02^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$ 35
- 16 Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence based on two lines (dashed line), and the independence based on the Matlab command *fminsearch* (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 0.02^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 6$ 35

FIGURE	Page
17	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values. 36
18	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 6$. The box has lines at the lower quartile, median, and upper quartile values. 36
19	Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence based on two lines (dashed line), and the independence based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$ 37
20	Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence based on two lines (dashed line), and the independence based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 6$ 37
21	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values. 38
22	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 6$. The box has lines at the lower quartile, median, and upper quartile values. 38

FIGURE	Page
23	Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 1^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$ 39
24	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 1^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values. 39
25	Histograms for the posterior distribution of $w_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000. 40
26	Histograms for the posterior distribution of $w_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000. 41
27	Histograms for the posterior distribution of $w_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 1^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000. 42
28	Posterior estimated mean functions for the Metropolis (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$ 44

29	Posterior estimated mean functions for the Metropolis (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 6$	44
30	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.	45
31	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 6$. The box has lines at the lower quartile, median, and upper quartile values.	45
32	Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the support adjusted and no shrinkage and the maximum resolution $m_0 = 5$	46
33	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the support adjusted and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.	46

FIGURE

Page

34 Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines(dashed line), and the independence sampler based on the Matlab command *fminsearch* (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 1^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$ 47

35 Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 1^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values. 47

36 Histograms for the posterior distribution of $w_{4,k}$ for $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 0.02^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000. 48

37 Histograms for the posterior distribution of $w_{4,k}$ for $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 0.5^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000. 49

38 Histograms for the posterior distribution of $w_{4,k}$ for $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 1^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000. 50

39 Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines(dashed line), and the independence sampler based on the Matlab command *fminsearch*(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 0.02^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$ 52

- 40 Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines(dashed line), and the independence sampler based on the Matlab command *fminsearch*(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 0.02^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 6$ 52
- 41 Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 0.02^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values. 53
- 42 Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 0.02^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 6$. The box has lines at the lower quartile, median, and upper quartile values. 53
- 43 Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines(dashed line), and the independence sampler based on the Matlab command *fminsearch*(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 0.5^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$ 54
- 44 Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 0.5^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values. 54
- 45 Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines(dashed line), and the independence sampler based on the Matlab command *fminsearch*(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in *Example 2* with $\sigma^2 = 1^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$ 55

FIGURE	Page
46	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 1^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values. 55
47	Histograms for the posterior probability of $s_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$, using the metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000. 56
48	Histograms for the posterior probability of $s_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$, using the metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000. 57
49	Histograms for the posterior probability of $s_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 1^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$, using the metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000. 58
50	Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence based on two lines(dashed line), and the independence based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$ 60
51	Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence based on two lines(dashed line), and the independence based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 6$ 60

FIGURE	Page
52	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values. 61
53	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 6$. The box has lines at the lower quartile, median, and upper quartile values. 61
54	Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence based on two lines(dashed line), and the independence based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$ 62
55	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values. 62
56	Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence based on two lines(dashed line), and the independence based on the Matlab command <i>fminsearch</i> (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 1^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$ 63
57	Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 1^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values. 63

FIGURE	Page
58	Histograms for the posterior distribution of $w_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.02^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000. 64
59	Histograms for the posterior distribution of $w_{4,k}$ for $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 0.5^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000. 65
60	Histograms for the posterior distribution of $w_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in <i>Example 2</i> with $\sigma^2 = 1^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000. 66
61	The estimated function and the boxplots of the direction parameter for the Metropolis algorithm and two independence samplers. The box has lines at the lower quartile, median, and upper quartile values. 68
62	Residuals for the Metropolis algorithm and two independence samplers. . . 69
63	Histograms of the coefficients $c_{0,k}$ for the Metropolis algorithm after the burn-in period. The scale on the y-axis is frequency with the maximum frequency 9000. 70
64	Histograms of the coefficients $c_{0,k}$ for the independence sampler based on two lines after the burn-in period. The scale on the y-axis is frequency with the maximum frequency 9000. 71
65	Histograms of the coefficients $c_{0,k}$ for the independence sampler based on the Matlab command <i>fminsearch</i> after the burn-in period. The scale on the y-axis is frequency with the maximum frequency 9000. 72

CHAPTER I

INTRODUCTION

1.1 Single index models

The classical regression model has the form,

$$y_i = X^i \beta + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

where $X^i = (x_{i1}, \dots, x_{ip})$, $\beta = (\beta_1, \dots, \beta_p)^T$, and ε_i is a random variable with zero mean and bounded variance, conditional on X^i . In a variety of fields there are data which cannot be analyzed with a straight line. For these data it is appropriate to allow for nonlinear regression models, which, however, have the drawback of the curse of dimensionality (Bellman 1961). Here we consider single index models (Härdle, *et al.*, 1993) which overcome the drawback and are useful in applications to econometrics and biometrics, where multidimensional regression models are often employed. Single index models are a special type of nonlinear regression models involving linear regression formulation (McCullagh and Nelder, 1983) and exploit the dependence of a scalar variable y_i upon a p -variate row vector X^i in the following form:

$$y_i = r(X^i \beta) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.2)$$

where $E(y_i|X^i) = r(X^i \beta)$, $Var(y_i|X^i) < \infty$, β is a p -column vector of unknown parameters, $\|\beta\| = 1$ and $r(\cdot)$ is an unknown function. The form (1.2) is equivalent to the traditional

This dissertation follows the style and format of the *Journal of the American Statistical Association*.

nonparametric regression where the scalar $X^i\beta$, called the *index* (Härdle, *et al.*, 1993), is given. Our goal is to estimate both β and $r(\cdot)$ in (1.2).

So far various methods have been developed in single index models. Among the methods are kernel smoothing (Härdle, *et al.*, 1993), local linear methods (Carroll, *et al.*, 1997), average derivatives (Stoker 1986; Härdle and Stoker 1989), and penalized splines (Yu and Ruppert 2002). Most of the methods assume that the function is smoothed. Here we introduce Bayesian methods for wavelet series that do not require strong assumptions on the function $r(\cdot)$.

1.2 Bayesian wavelet methods for single index models

(Non)parametric methods have been proposed for smooth functions $r(\cdot)$, but appear to be inappropriate for discontinuities and functions with spikes. We consider a wavelet series which provides a good approximation to any function in L^2 space (1.1 in Appendix A).

The wavelet series consists of two basis functions, a scaling function $\phi(x)$ and a wavelet function $\psi(x)$. Translations and dilations of these two basis functions, $\{\phi(2^jx - k)\}$ and $\{\psi(2^jx - k)\}$ with j, k integers, are formed to represent the function $r(\cdot)$. Wavelet series have the advantage of being localized.

Most of the wavelet contributions in the literature assume equally spaced data. In the model (1.2), the indices are not equally spaced for all directions β . A possible solution to this problem is to replace the unequally spaced data by equally spaced data and proceed as in the equally spaced case. If there are only a few unequally spaced indices, removing these indices to make the data equally spaced lead to efficient implementations (Antoniadis, Gregoire and McKeague, 1994; Cai and Brown, 1997; Hall and Turlach, 1997; Sardy *et al.*, 1997).

We consider Bayesian methods, which require prior models for estimating the direction parameter β and the function. Two possible ways for representing the function based

on a wavelet series are proposed: no shrinkage means all coefficients of a wavelet series are used and shrinkage means some coefficients are set to zero. In both of these the prior models used for non-zero scaling(wavelet) coefficients are hierarchical, and since shrinkage rules allow some wavelet coefficients to be zero, we propose using "pseudo prior" models for the vanishing coefficients (Müller and Vidakovic, 1999). For details, see section 3.2.

We transform the p -dimensional direction parameter β into the spherical polar coordinates parameter θ , since this is a convenient way to impose the necessary constraint of $\|\beta\| = 1$. The posterior distribution of the direction parameter is not known and we employ a Metropolis algorithm and an independence sampler, which require proposal distributions. A normal distribution is proposed as the proposal distribution for the direction parameter θ . Its mode(mean) and/or variance may need to be chosen carefully. We suggest some simple approaches for selecting these two values.

1.3 Outline of the dissertation

Chapter II describes Bayesian methods for the wavelet series with a single covariate. Chapter III presents applications of wavelet-based Bayesian methods in the single index model. Chapter IV presents simulation studies and an application to real data. Finally, Chapter V discusses results and further work.

CHAPTER II

BAYESIAN METHODS FOR WAVELETS

For several decades, wavelet analysis has been developed and applied to many fields such as data compression, signal and image processing, numerical analysis, chemistry, and astronomy. In nonparametric regression problems, wavelets are used to estimate an unknown function $r(x)$ under weak assumptions about the nature of $r(x)$. The following sections describe how to represent a function with a wavelet series and how to perform posterior inference on the parameters of the wavelet series in the case of unevenly spaced data. For details about wavelet theory see Daubechies (1992), Chui (1992), and Meyer (1992).

2.1 Wavelet series

The *scaling* function $\phi(x)$ and the wavelet $\psi(x)$ play an important role in constructing a wavelet series. They describe the smooth parts and the detail parts of the function $r(x)$, respectively. The generalized forms of ϕ and ψ use a dilation parameter j and a translation parameter k as

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^jx - k), \quad \psi_{j,k}(x) = 2^{j/2}\psi(2^jx - k). \quad (2.1)$$

The collections $\{\phi_{j,k}, \psi_{j,k}\}$ are an orthogonal basis:

$$\begin{aligned} \int \phi_{j,k}(x)\phi_{j',k'}(x)dx &= \delta_{k,k'} \\ \int \psi_{j,k}(x)\phi_{j',k'}(x)dx &= 0 \\ \int \psi_{j,k}(x)\psi_{j',k'}(x)dx &= \delta_{j,j'}\delta_{k,k'} \end{aligned}$$

where

$$\delta_{i,j} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

An orthogonal wavelet series is the following linear representation of the square integrable function $r(x)$:

$$r(x; m_0) = \sum_{k \in \mathbf{Z}} c_{J_0, k} \phi_{J_0, k}(x) + \sum_{j=J_0}^{m_0} \sum_{k \in \mathbf{Z}} w_{j, k} \psi_{j, k}(x), \quad (2.2)$$

for any integer J_0 where $c_{J_0, k} = \int r(x) \phi_{J_0, k}(x) dx$ and $w_{j, k} = \int r(x) \psi_{j, k}(x) dx$. The function $r(\cdot; m_0)$ converges in mean square to r as $m_0 \rightarrow \infty$. In the representation (2.2) we will assume $J_0 = 0$ (Müller and Vidakovic, 1999).

The approximation (2.2) is analogous to an orthogonal Fourier series. However, wavelet series are localized both in time and frequency, while the Fourier functions are localized in frequency only. In addition, localization properties allow the representation (2.2) to be parsimonious.

2.2 Bayesian approaches to wavelet series

A wavelet series provides a good approximation to any square integrable function. We consider a Bayesian approach to wavelet series. To implement the posterior inference, prior models for estimating the function $r(x)$ are required. Adaptive shrinkage (Chipman, Kolaczyk and McCulloch, 1997), multiple shrinkage (Clyde, Parmigiani and Vidakovic, 1998), and nonlinear shrinkage (Vidakovic, 1998) are Bayesian methods for wavelet series with the assumption of equally spaced data.

Since localization allows for a parsimonious representation, we can consider two ways to represent the wavelet series of the function $r(x)$. First, all wavelet coefficients are kept in the wavelet representation. Second, wavelet shrinkage rules are proposed to obtain model parsimony. For both representations, since a coefficient has information about the features of the function at a certain scale and location, a prior model for a non-zero scaling(wavelet)

coefficient should depend on its resolution level. For example, a normal prior with variance dependent on a resolution level may be used. In wavelet shrinkage rules we follow Carlin and Chib (1995) in using pseudo-priors for zero wavelet coefficients as explained in subsection (3.2.2)

For equally spaced data, the coefficients are usually assumed to be a priori independent (Yan and Kohn, 1999).

CHAPTER III

BAYESIAN METHODS FOR WAVELET REPRESENTATIONS OF SINGLE INDEX
MODELS**3.1 Wavelet representations of single index models**

We consider a wavelet series which is an approximation to the function r in single index models. From (2.2) when a covariate x is replaced by a single index $X^i\beta$, a wavelet series in single index models is

$$r(X^i\beta) = \sum_{k \in \mathbb{Z}} c_{0,k} \phi_{0,k}(X^i\beta) + \sum_{j=0}^{m_0} \sum_{k \in \mathbb{Z}} w_{j,k} \psi_{j,k}(X^i\beta), \quad (3.1)$$

where m_0 is given.

Shrinkage rules can be implemented to take advantage of the parsimony of the wavelet representation. To apply these to the wavelet series (3.1), we use indicators $s_{j,k}$ defined by

$$s_{j,k} = \begin{cases} 1, & \text{if } w_{j,k} \text{ is included in the wavelet series,} \\ 0, & \text{if } w_{j,k} \text{ is not included in the wavelet series.} \end{cases}$$

We define a wavelet shrinkage model as

$$r(X^i\beta) = \sum_{k \in \mathbb{Z}} c_{0,k} \cdot \phi_{0,k}(X^i\beta) + \sum_{j=0}^{m_0} \sum_{k \in \mathbb{Z}} s_{j,k} \cdot w_{j,k} \cdot \psi_{j,k}(X^i\beta), \quad (3.2)$$

where m_0 is given.

3.2 Likelihood function and prior distributions

To apply Bayesian methods we need the likelihood function and the prior of all parameters.

From (1.2), (3.1), (3.2) and assuming normality of the errors, the likelihood function is

$$P(Y|\sigma^2, \{c_{0,k}\}, \{w_{j,k}\}, \beta, \mathbf{X}, m_0) = \prod_{i=1}^n P(y_i|\sigma^2, \{c_{0,k}\}, \{w_{j,k}\}, \beta, X^i, m_0)$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \mathbf{Q}(y_i)\right), \quad (3.3)$$

where for the no shrinkage rule

$$\mathbf{Q}(y_i) = \mathcal{Q}_1(y_i) = \left[y_i - \sum_{k \in \mathbb{Z}} c_{0,k} \phi_{0,k}(X^i \boldsymbol{\beta}) - \sum_{j=0}^{m_0} \sum_{k \in \mathbb{Z}} w_{j,k} \psi_{j,k}(X^i \boldsymbol{\beta}) \right]^2, \quad (3.4)$$

or for shrinkage rules

$$\mathbf{Q}(y_i) = \mathcal{Q}_2(y_i) = \left[y_i - \sum_{k \in \mathbb{Z}} c_{0,k} \cdot \phi_{0,k}(X^i \boldsymbol{\beta}) - \sum_{j=0}^{m_0} \sum_{k \in \mathbb{Z}} s_{j,k} \cdot w_{j,k} \cdot \psi_{j,k}(X^i \boldsymbol{\beta}) \right]^2. \quad (3.5)$$

In single index models what we are interested in are the function r and the direction $\boldsymbol{\beta}$ in (1.2). In particular, we can consider the direction as points on the circumference of a unit circle in two dimensions, on the surface of a unit sphere in three dimensions and so on. The unit vector $\boldsymbol{\beta}$ is transformed to the surface of a $(p-1)$ -dimensional hypersphere S_{p-1} of unit radius and having its center at the origin.

In general, it is convenient to transform $\boldsymbol{\beta}$ to a spherical polar coordinate $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{p-1})^T$, with

$$t_d(\boldsymbol{\theta}) = \beta_d = \sin(\theta_{d-1}) \prod_{j=0}^{p-d} \cos(\theta_{p-j}), \quad d = 1, \dots, p, \quad (3.6)$$

$\sin(\theta_0) = \cos(\theta_p) = 1$, $0 < \theta_1 < 2\pi$, and $-\frac{\pi}{2} < \theta_l < \frac{\pi}{2}$, $l = 2, \dots, p-1$ (see Figures 1 and 2).

Using the spherical polar coordinate $\boldsymbol{\theta}$ instead of the direction $\boldsymbol{\beta}$, i.e., $X^i \boldsymbol{\beta} = X^i \mathbf{T}(\boldsymbol{\theta})$ where $\boldsymbol{\beta} = \mathbf{T}(\boldsymbol{\theta}) = (t_1(\boldsymbol{\theta}), \dots, t_p(\boldsymbol{\theta}))^T$, we rewrite the likelihood function:

$$P(Y|\sigma^2, \{c_{0,k}\}, \{w_{j,k}\}, \boldsymbol{\theta}, \mathbf{X}, m_0) = P(Y|\sigma^2, \{c_{0,k}\}, \{w_{j,k}\}, \boldsymbol{\beta}, \mathbf{X}, m_0), \quad (3.7)$$

where for the no shrinkage rule

$$\mathbf{Q}(y_i) = \mathcal{Q}_3(y_i) = \left[y_i - \sum_{k \in \mathbb{Z}} c_{0,k} \phi_{0,k}(X^i \mathbf{T}(\boldsymbol{\theta})) - \sum_{j=0}^{m_0} \sum_{k \in \mathbb{Z}} w_{j,k} \psi_{j,k}(X^i \mathbf{T}(\boldsymbol{\theta})) \right]^2, \quad (3.8)$$

or for shrinkage rules

$$\mathbf{Q}(y_i) = \mathcal{Q}_4(y_i) = \left[y_i - \sum_{k \in \mathbb{Z}} c_{0,k} \cdot \phi_{0,k}(X^i \mathbf{T}(\boldsymbol{\theta})) - \sum_{j=0}^{m_0} \sum_{k \in \mathbb{Z}} s_{j,k} \cdot w_{j,k} \cdot \psi_{j,k}(X^i \mathbf{T}(\boldsymbol{\theta})) \right]^2. \quad (3.9)$$

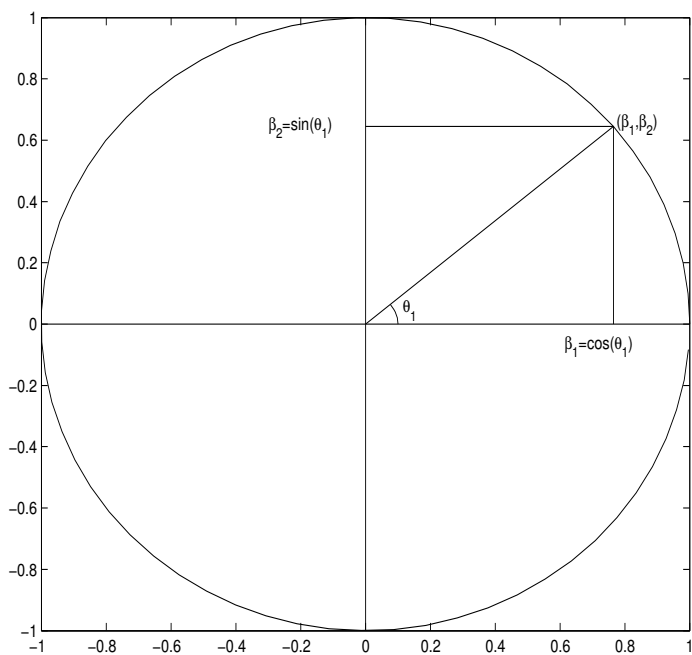


Figure 1. Circular polar coordinates.

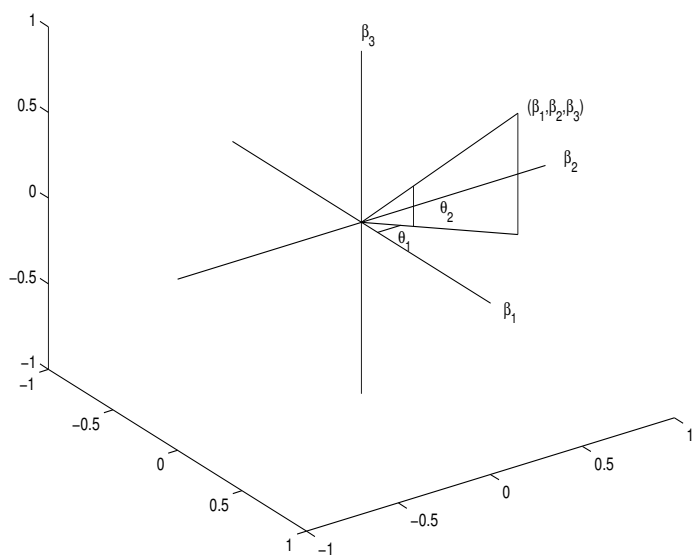


Figure 2. Spherical polar coordinates.

3.2.1 Prior models for fixed resolution

To implement posterior inference, we require prior models from (3.3), (3.4), (3.6), and (3.7), given the resolution m_0 . For notational simplicity, let ' \sim ' mean 'distributed as', let $N(\mu, \eta)$ denote a normal distribution with mean μ and variance η , let $IG(a, b)$ denote an inverse gamma distribution with mean $\frac{1}{(a-1)b}$ and variance $\frac{1}{(a-1)^2(a-2)b^2}$, and let $P(\cdot)$ generically denote a probability density function. We define a hierarchical prior model on each coefficient $c_{0,k}$ or $w_{j,k}$. We assume a normal distribution as a prior model for each coefficient with a variance depending on the resolution level and a hyperparameter τ

$$c_{0,k}|\tau \sim N(0, \tau), \quad w_{j,k}|\tau \sim N(0, \tau 2^{-j}). \quad (3.10)$$

The scale factor 2^{-j} is due to the factor $2^{j/2}$ in the definition of $\psi_{j,k}(x)$ from (2.1). For details, see Müller and Vidakovic (1998). We define prior models for σ^2 and τ as

$$\tau \sim IG(a_\tau, b_\tau), \quad \sigma^2 \sim IG(a_v, b_v). \quad (3.11)$$

A prior for the direction θ is uniform

$$P(\theta) = \frac{1}{2\pi} \left(\frac{1}{\pi}\right)^{p-2}, \quad (3.12)$$

where θ is $(p-1) \times 1$ and $0 < \theta_1 < 2\pi, -\frac{\pi}{2} < \theta_d < \frac{\pi}{2}, \quad d = 2, \dots, p-1$.

Let $\Omega_1 = [\{c_{0,k}\}, \{w_{j,k}\}, \tau, \sigma^2, \theta]$ and let \propto mean 'proportional to'. From (3.10)-(3.12), the joint prior distribution is

$$P(\Omega_1) = \prod_k P(c_{0,k}|\tau) \cdot \prod_{j,k} P(w_{j,k}|\tau) \cdot P(\tau|a_\tau, b_\tau) \cdot P(\sigma^2|a_v, b_v) \cdot P(\theta). \quad (3.13)$$

3.2.2 Mixture priors for wavelet shrinkage

The wavelet series is a parsimonious representation, i.e., some coefficients at high levels of detail are close to zero. To allow for wavelet shrinkage we need a prior model for a

wavelet coefficient which is a hierarchical mixture model with a point-mass at zero and a continuous distribution for non-zero values. We propose a prior for the indicator $s_{j,k}$, i.e., $j > 0$

$$\chi_j = Pr(s_{j,k} = 1) = \alpha^j, \quad (3.14)$$

where Pr means 'probability'.

Our prior for a wavelet coefficient is defined as

$$w_{j,k} | \tau, s_{j,k} = 1 \sim N(0, \tau 2^{-j}), \quad (3.15)$$

$$w_{j,k} | \tau, s_{j,k} = 0 \sim h(w_{j,k}), \quad (3.16)$$

the prior for a scaling coefficient is as in (3.10), and choices of $h(w_{j,k})$ will be discussed below.

In words, wavelet coefficients which are non-zero have decreasing probabilities χ_j at higher resolution levels. If $s_{j,k} = 0$, then $w_{j,k}$ is not included in the orthogonal wavelet series and in the likelihood function. To achieve posterior inference for the zero coefficients we need to assume a "pseudo-prior" $h(w_{j,k})$, as described below.

Shrinkage rules make small coefficients shrink significantly more than large values, because the (true) fine detail coefficients tend to be small, i.e., high level coefficients are shrunk more strongly than low level coefficients. Thus, the prior probability χ_j and the scale factor 2^{-j} have information on the rate of decay in the magnitude of the wavelet coefficients and indirectly on the smoothness of the function r . Other methods of controlling the rate of decay are possible. For example, Abramovich, Sapatinas, and Silverman (1998) used $\chi_j = \min[1, C_2(1/2)^{\lambda_1 j}]$ and $\tau_j = C_1(1/2)^{\lambda_2 j}$ with C_1 and C_2 determined in an empirical Bayes fashion, where λ_1 and λ_2 are smoothness indices. Chipman, Kolaczyk, and McCulloch (1997) took $\chi_j \propto f_j(1/2)^j$, where f_j is the fraction of empirical wavelet coefficients larger than a certain cutoff.

A variable number of coefficients is implied by the prior model. This leads to a variable dimension problem. To overcome this problem Carlin and Chib (1995) proposed to set up artificial priors, i.e., "pseudo priors" on coefficients which are not currently included in the likelihood. The following pseudo prior $h(w_{j,k})$ is proposed:

$$w_{j,k}|s_{j,k} = 0 \sim N(\hat{w}_{j,k}, \hat{\sigma}_{j,k}^2), \quad (3.17)$$

where $\hat{w}_{j,k}$ and $\hat{\sigma}_{j,k}^2$ are some initial guesses of the marginal posterior mean and variance of $w_{j,k}$, respectively. Choosing $\hat{w}_{j,k}$ and $\hat{\sigma}_{j,k}^2$ is discussed in Müller and Vidakovic (1999). In this paper, after an initial burn-in period of T_0 , the values of $\hat{w}_{j,k}$ and $\hat{\sigma}_{j,k}^2$ are computed by the ergodic mean and variance of $w_{j,k}$ over the initial burn-in period.

To achieve the efficient implementation of MCMC schemes, it is important to have a good choice of the pseudo prior $h(w_{j,k})$, which changes the simulated Markov chain, but not the asymptotic distribution. The asymptotic distribution is always dominated by the posterior distribution, which is derived from the likelihood and the prior.

We complete the model with a hyperprior for α in (3.14):

$$\alpha \sim Beta(a_\alpha, b_\alpha). \quad (3.18)$$

Let $\Omega_2 = [\{c_{0,k}\}, \{s_{j,k}\}, \{w_{j,k}\}, \alpha, \tau, \sigma^2, \theta]$. From (3.11), (3.12), and (3.14)-(3.18), the joint prior is

$$\begin{aligned} P(\Omega_2) &= \prod_k P(c_{0,k}|\tau) \cdot \prod_{j,k} P(s_{j,k}|\alpha) \cdot \prod_{s_{j,k} \in \{1\}} P(w_{j,k}|s_{j,k} = 1, \tau) \\ &\quad \times \prod_{s_{j,k} \in \{0\}} P(w_{j,k}|s_{j,k} = 0) \cdot P(\alpha|a_\alpha, b_\alpha) \cdot P(\tau|a_\tau, b_\tau) \cdot P(\sigma^2|a_\nu, b_\nu) \\ &\quad \times P(\theta) \\ &\propto \prod_k P(c_{0,k}|\tau) \cdot \prod_{j,k} P(s_{j,k}) \cdot \prod_{s_{j,k} \in \{1\}} P(w_{j,k}|s_{j,k} = 1, \tau) \\ &\quad \times \prod_{s_{j,k} \in \{0\}} P(w_{j,k}|s_{j,k} = 0) \cdot P(\alpha|a_\alpha, b_\alpha) \cdot P(\tau|a_\tau, b_\tau) \cdot P(\sigma^2|a_\nu, b_\nu). \end{aligned} \quad (3.19)$$

3.3 Posterior inference and MCMC schemes

We derive full conditional distributions to implement an MCMC scheme. Let $Z = \mathbf{X}\beta \in \{z_{min}, z_{max}\}$ which are computed based on the maximum(minimum) of $X\beta$ (see Appendix A) and let the boundaries of the translations k of the coefficients $\{c_{0,k}, w_{j,k}\}$ be denoted $[a_0, b_0]$ and $[a_j, b_j]$, $j = 0, \dots, m_0$, respectively, (see Subsection 3.5.1).

3.3.1 Posterior inference for no shrinkage

Recall $\Omega_1 = [\{c_{0,k}\}, \{w_{j,k}\}, \tau, \sigma^2, \theta]$ and let $\Omega_1(-\xi)$ denote Ω_1 without the parameter ξ which could be any of $\{c_{0,k}\}, \{w_{j,k}\}, \tau, \sigma^2$, or θ . We show the following calculations in Appendix A. From (3.7), (3.8), and (3.13), the joint distribution of the data Y and the parameters Ω_1 , conditional on \mathbf{X} and the resolution m_0 is

$$P(Y, \Omega_1 | \mathbf{X}, m_0) = P(Y | \sigma^2, \{c_{0,k}\}, \{w_{j,k}\}, \theta, \mathbf{X}, m_0) \cdot P(\Omega_1). \quad (3.20)$$

Then the posterior distribution of σ^2 , given $\Omega_1(-\sigma^2), Y, \mathbf{X}$ and m_0 , is

$$\sigma^2 | \Omega_1(-\sigma^2), Y, \mathbf{X}, m_0 \sim IG \left[\frac{n}{2} + a_v, \left(\frac{1}{b_v} + \frac{1}{2} \sum_{i=1}^n Q_3(y_i) \right)^{-1} \right]. \quad (3.21)$$

The posterior distribution of τ , given $\Omega_1(-\tau), Y, \mathbf{X}$ and m_0 , is

$$\tau | \Omega_1(-\tau), Y, \mathbf{X}, m_0 \sim IG \left[\frac{s}{2} + a_\tau, \left(\frac{1}{b_\tau} + \frac{1}{2} \sum_{a_0}^{b_0} c_{0,k}^2 + \frac{1}{2} \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} \frac{w_{j,k}^2}{2^{-j}} \right) \right], \quad (3.22)$$

where $s = b_0 - a_0 + 1 + \sum_{j=0}^{m_0} (b_j - a_j + 1)$.

The posterior distribution of c_{0,k_1} , given $\Omega_1(-c_{0,k_1}), Y, \mathbf{X}$ and m_0 , is

$$c_{0,k_1} | \Omega_1(-c_{0,k_1}), Y, \mathbf{X}, m_0 \sim N(\mu_{k_1}, \sigma_{k_1}^2), \quad (3.23)$$

where μ_{k_1} and $\sigma_{k_1}^2$ are given in Appendix B.

The posterior distribution of w_{j_2, k_2} , given $\Omega_1(-w_{j_2, k_2}), Y, \mathbf{X}$ and m_0 , is

$$w_{j_2, k_2} | \Omega_1(-w_{j_2, k_2}), Y, \mathbf{X}, m_0 \sim N(\mu_{(j_2, k_2)}, \sigma_{(j_2, k_2)}^2), \quad (3.24)$$

where $\mu_{(j_2, k_2)}$ and $\sigma_{(j_2, k_2)}^2$ are given in Appendix B.

The posterior distribution of θ , given $\Omega_1(-\theta), Y, \mathbf{X}$ and m_0 , cannot be determined explicitly. See B-6 in Appendix B.

3.3.2 Posterior inference for wavelet shrinkage

Recall that $\Omega_2 = [\{c_{0,k}\}, \{s_{j,k}\}, \{w_{j,k}\}, \alpha, \tau, \sigma^2, \theta]$. We show the following calculations in Appendix C. From (3.7), (3.9), and (3.19), the joint distribution of the data Y and the parameters, conditional on X and the resolution m_0 , is

$$P(Y, \Omega_2 | \mathbf{X}, m_0) = P(Y | \sigma^2, \{c_{0,k}\}, \{s_{j,k}\}, \{w_{j,k}\}, \theta, \mathbf{X}, m_0) \cdot P(\Omega_2). \quad (3.25)$$

Then the posterior distribution of σ^2 , given $\Omega_2(-\sigma^2), Y, \mathbf{X}$ and m_0 , is

$$\sigma^2 | \Omega_2(-\sigma^2), Y, \mathbf{X}, m_0 \sim IG \left[\frac{n}{2} + a_v, \left(\frac{1}{b_v} + \frac{1}{2} \sum_{i=1}^n Q_4(y_i) \right)^{-1} \right]. \quad (3.26)$$

The posterior distribution of τ , given $\Omega_1(-\tau), Y, \mathbf{X}$ and m_0 , is

$$\tau | \Omega_2(-\tau), Y, \mathbf{X}, m_0 \sim IG \left[\frac{s}{2} + a_\tau, \left(\frac{1}{b_\tau} + \frac{1}{2} \sum_{k=a_0}^{b_0} c_{0,k}^2 + \frac{1}{2} \sum_{j=0}^{m_0} \sum_{\substack{k=a_j \\ j,k \in \{s_{j,k}=1\}}}^{b_j} \frac{w_{j,k}^2}{2^{-j}} \right) \right], \quad (3.27)$$

where $s = b_0 - a_0 + 1 + \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} s_{j,k}$.

The posterior distribution of c_{0,k_1} , given $\Omega_2(-c_{0,k_1}), Y, \mathbf{X}$ and m_0 , is

$$c_{0,k_1} | \Omega_2(-c_{0,k_1}), Y, \mathbf{X}, m_0 \sim N(\mu_{k_1}, \sigma_{k_1}^2), \quad (3.28)$$

where μ_{k_1} and $\sigma_{k_1}^2$ are given in Appendix C.

The posterior distribution of w_{j_2, k_2} , given $\Omega_2(-w_{j_2, k_2}), s_{j_2, k_2} = 1, Y, \mathbf{X}$ and m_0 , is

$$w_{j_2, k_2} | \Omega_2(-w_{j_2, k_2}), s_{j_2, k_2} = 1, Y, \mathbf{X}, m_0 \sim N(\mu_{(j_2, k_2)}, \sigma_{(j_2, k_2)}^2), \quad (3.29)$$

where $\mu_{(j_2, k_2)}$ and $\sigma_{(j_2, k_2)}^2$ are in Appendix C.

The posterior distribution of w_{j_2, k_2} , given $\Omega_2(-w_{j_2, k_2}), s_{j_2, k_2} = 0, Y, \mathbf{X}$ and m_0 , is

$$w_{j_2, k_2} | \Omega_2(-w_{j_2, k_2}), s_{j_2, k_2} = 0, Y, X, m_0 \sim N(\hat{w}_{j, k}, \hat{\sigma}_{j, k}^2), \quad (3.30)$$

where $\hat{w}_{j, k}, \hat{\sigma}_{j, k}^2$ are initial guesses of the marginal posterior mean and variance of $w_{j, k}$.

The posterior distribution of s_{j_2, k_2} , given $\Omega_2(-s_{j_2, k_2}), Y, \mathbf{X}$ and m_0 , is

$$s_{j_2, k_2} | \Omega_2(-s_{j_2, k_2}), Y, \mathbf{X} \sim \begin{cases} Pr(s_{j, k} = 0) \propto pr_0 = \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(Y_i)\right) (1 - \alpha^{j_2}) h(w_{j_2, k_2}), \\ Pr(s_{j, k} = 1) \propto pr_1 = \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(Y_i)\right) (\alpha^{j_2}) P(w_{j_2, k_2} | s_{j_2, k_2} = 1, \tau), \end{cases} \quad (3.31)$$

with probabilities $Pr(s_{j, k} = 0) = pr_0 / (pr_0 + pr_1)$ and $Pr(s_{j, k} = 1) = pr_1 / (pr_0 + pr_1)$.

The posterior distributions of α and θ , given $\Omega_2(-\alpha, -\theta), Y, \mathbf{X}$ and m_0 , cannot be determined explicitly (see Appendix C).

3.3.3 MCMC schemes

We describe MCMC schemes for no shrinkage and shrinkage from Subsection 3.3.1 and Subsection 3.3.2. Starting from initial values for the parameters Ω_1 or Ω_2 , the parameters

are updated, one at time. For details about the initialization see Subsection 3.5.2. The posterior distributions for each step are calculated in Appendix B and Appendix C.

The graphical model of the single index model with no shrinkage is shown in Figure 3. For no shrinkage, the MCMC scheme is as follows.

- 1 To update σ^2 , generate σ^2 from the complete inverse gamma conditional distribution.
- 2 To update τ , generate τ from the complete inverse gamma conditional distribution.
- 3 To update c_{0,k_1} , generate c_{0,k_1} from the complete normal conditional distribution.
- 4 To update w_{j_2,k_2} , generate w_{j_2,k_2} from the complete normal conditional distribution.
- 5a Generate $\tilde{\theta}$ from a normal proposal distribution with mean θ and variance $\hat{\sigma}_1^2$, and compute the acceptance probability

$$A(\theta, \tilde{\theta}) = \min \left(1, \frac{P(\tilde{\theta} | \Omega_1(-\theta))}{P(\theta | \Omega_1(-\theta))} \right).$$

- 5b Generate $\tilde{\theta}$ from a normal proposal distribution with mean μ_θ and variance $\hat{\sigma}_2^2$, and compute the acceptance probability

$$A(\theta, \tilde{\theta}) = \min \left(1, \frac{P(\tilde{\theta} | \Omega_1(-\theta)) \cdot q(\theta | \mu_\theta, \hat{\sigma}_2^2)}{P(\theta | \Omega_1(-\theta)) \cdot q(\tilde{\theta} | \mu_\theta, \hat{\sigma}_2^2)} \right).$$

Steps 1-4 are Gibbs sampling steps, that is, the parameters are updated by generations from the complete conditional distributions. Steps 5a and 5b are a Metropolis step (Metropolis *et al.*, 1953) and an independence sampler step (Tierney, 1994), respectively. The values μ_θ , $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ are described below. We use a short primary run to find a suitable initial value for the direction parameter.

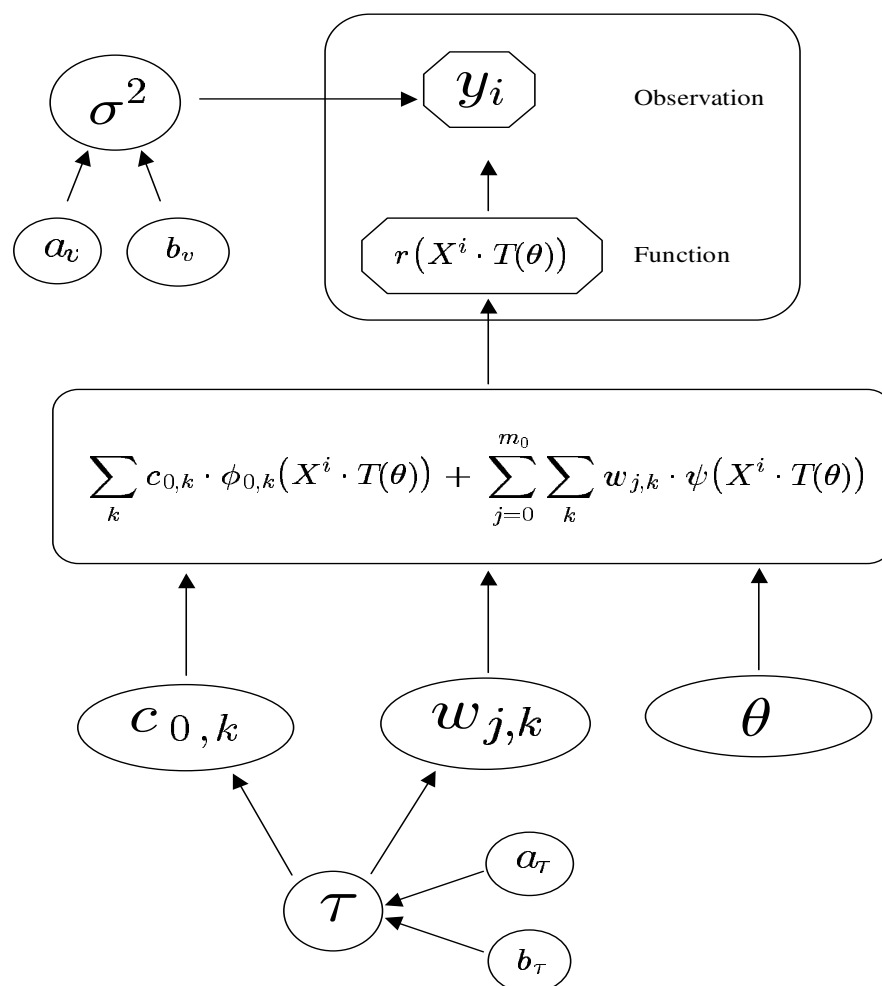


Figure 3. A single index model without shrinkage as a graphical model.

The graphical model of the single index model with shrinkage is shown in Figure 4. For shrinkage, the MCMC scheme uses the following steps.

- 1 To update σ^2 , generate σ^2 from the complete inverse gamma conditional distribution.
- 2 To update τ , generate τ from the complete inverse gamma conditional distribution.
- 3 To update c_{0,k_1} , generate c_{0,k_1} from the complete normal conditional distribution.
- 4a To update w_{j_2,k_2} , given $s_{j_2,k_2} = 1$, generate w_{j_2,k_2} from the complete normal conditional distribution.
- 4b To update w_{j_2,k_2} , given $s_{j_2,k_2} = 0$, generate w_{j_2,k_2} from the complete normal conditional distribution.
- 5 To update each indicator s_{j_2,k_2} , generate the complete conditional posterior probability.

- 6 Generate $\alpha_2 \sim N(\alpha_1, 0.1^2)$, compute

$$A(\alpha_1, \alpha_2) = \min \left(1, \frac{\alpha_2^{a\alpha} (1 - \alpha_2)^{b\alpha}}{\alpha_1^{a\alpha} (1 - \alpha_1)^{b\alpha}} \cdot \prod_{j,k} \left[\frac{\alpha_2^j}{\alpha_1^j} \right]^{s_{j,k}} \left[\frac{1 - \alpha_2^j}{1 - \alpha_1^j} \right]^{1 - s_{j,k}} \right),$$

and with probability $A(\alpha_1, \alpha_2)$ replace α_1 by α_2 . Otherwise leave α_1 unchanged.

- 7a Generate $\tilde{\theta}$ from a normal proposal distribution with mean θ and variance $\hat{\sigma}_1^2$, and compute the acceptance probability

$$A(\theta, \tilde{\theta}) = \min \left(1, \frac{P(\tilde{\theta}|\Omega_1(-\theta))}{P(\theta|\Omega_1(-\theta))} \right).$$

- 7b Generate $\tilde{\theta}$ from a normal proposal distribution with mean μ_θ and variance $\hat{\sigma}_2^2$, and compute the acceptance probability

$$A(\theta, \tilde{\theta}) = \min \left(1, \frac{P(\tilde{\theta}|\Omega_1(-\theta)) \cdot q(\theta|\mu_\theta, \hat{\sigma}_2^2)}{P(\theta|\Omega_1(-\theta)) \cdot q(\tilde{\theta}|\mu_\theta, \hat{\sigma}_2^2)} \right).$$

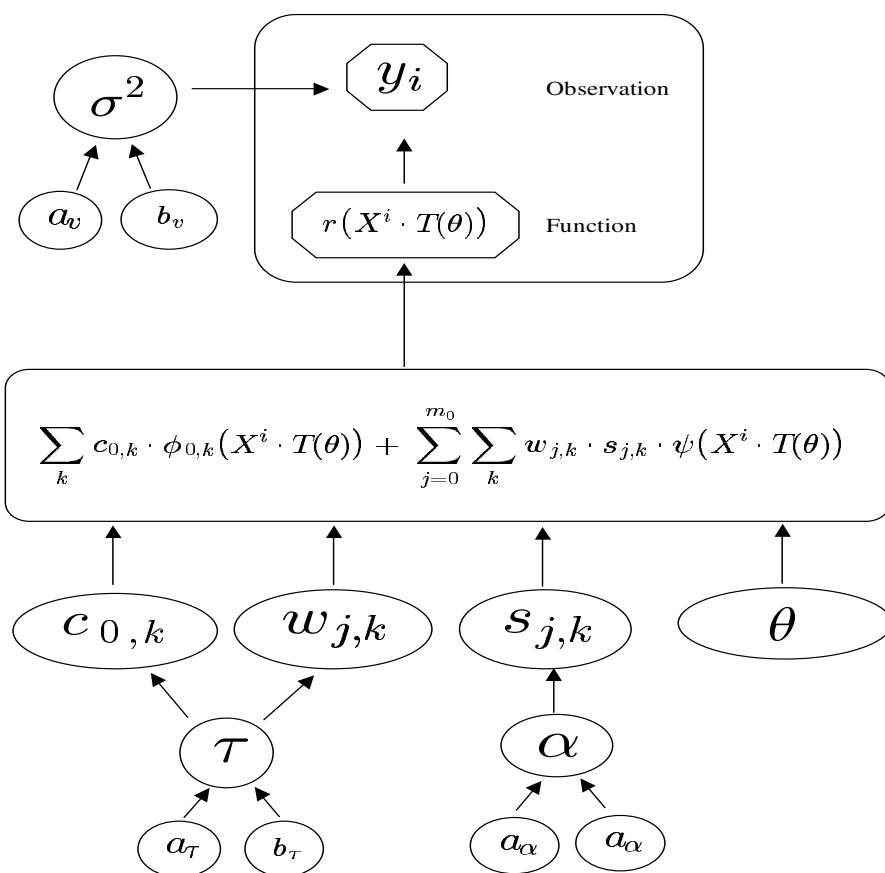


Figure 4. A single index model with shrinkage as a graphical model.

Steps 1-5 are Gibbs sampling steps, that is, the parameters are updated by generations from the complete conditional distributions. Steps 6 and 7a(b) are a Metropolis step (Metropolis *et al.*, 1953) or an independence sampler step (see Subsection 5.3.4 in Gilks, *et al.*, 1996), respectively. We use a short primary run to find an initial direction with all $s_{j,k}$'s equal to 1. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are computed based on acceptance rates (see Subsection 3.5.2), and for μ_θ see Section 3.4.

3.4 An independence sampler for a normal proposal distribution

We propose a normal distribution as the proposal distribution for the direction θ . When an independence sampler is used to generate samples of the direction θ from a normal distribution, we need to specify its mode (mean). Here we introduce two ways to find the mode (mean). Suppose that the target distribution of the direction θ is close to a normal distribution (B-6 in Appendix B). The mode (mean) is a value which maximizes the posterior distribution of the direction θ . We can use the Matlab function *fminsearch* which implements an unconstrained nonlinear optimization method, that is, a direct search method that does not use numerical or analytic gradients. A limitation of *fminsearch* is that it often can not handle a discontinuity, particularly if the discontinuity occurs near the solution. A new approach to search for a mode (mean) is introduced as follows.

Suppose that the posterior distribution is unimodal. The steps to approximate a mode (mean) for the posterior distribution are shown schematically in Figure (5). Suppose the MCMC algorithm is being performed, and let p_1 be the current value of the direction parameter and $\hat{\sigma}_\theta^2$ the variance of the proposal distribution. Let $P(\theta)$ denote the (known) function that is proportional to the full conditional of θ . The slope of the line connecting two points $(\theta_0, P(\theta_0))$, $(\theta_1, P(\theta_1))$ is

$$\gamma = \frac{P(\theta_1) - P(\theta_0)}{\theta_1 - \theta_0}.$$

- Step 1 - To see where the current p_1 is relative to the mode, compare two slopes from 3 points: $(p_1 - \hat{\sigma}_\theta, P(p_1 - \hat{\sigma}_\theta))$, $(p_1, P(p_1))$, and $(p_1 + \hat{\sigma}_\theta, P(p_1 + \hat{\sigma}_\theta))$. If the slopes are negative, the current direction is to the right of the mode (mean). On the other hand, if the slopes are positive, the current direction is to the left of the mode (mean).
- Step 2 - If the slopes are negative(positive), shift the previous p_1 to left(right) $p_2 = p_1 - \hat{\sigma}_\theta$ ($p_2 = p_1 + \hat{\sigma}_\theta$), repeat Step 1 and shift until the signs of the two slopes are different.
- Step 3 - Consider the three points $(L, P(L))$, $(p_3, P(p_3))$, and $(R, P(R))$, where $L = p_3 - \hat{\sigma}_\theta$, p_3 , $R = p_3 + \hat{\sigma}_\theta$, and the signs of the two slopes are different.
- Step 4 - When $P(L)$ is greater(less) than $P(R)$, the middle point p_3 is on the right side(left side) of the true mode.
- Step 5 - Calculate the slope γ of the line corresponding to the two points that are on the same side of the mode. Construct a line passing through the point on the opposite side of the mode and having slope $-\gamma$.
- Step 6 - Define a candidate mode (mean) to be the abscissa at which the two lines intersect.

3.5 Computational issues

3.5.1 Range of translation parameters

We consider minimum phase Daubechies wavelets with compact support which ensure a finite range of values. The supports of the scaling function $\phi(x)$ and the wavelet function $\psi(x)$ are $[0, 2N - 1]$ and $[-N, N - 1]$, respectively, where N is the number of vanishing moments. We can easily calculate the supports of $\phi_{j,k}(x)$, $\psi_{j,k}(x)$ and the range of the translation parameter k for $\phi_{j,k}(x)$, $\psi_{j,k}(x)$. The supports of $\phi_{j,k}(x)$ and $\psi_{j,k}(x)$ are

$$\left[\frac{k}{2^j}, \frac{2N - 1 + k}{2^j} \right], \quad \left[\frac{1 - N + k}{2^j}, \frac{N + k}{2^j} \right]. \quad (3.32)$$

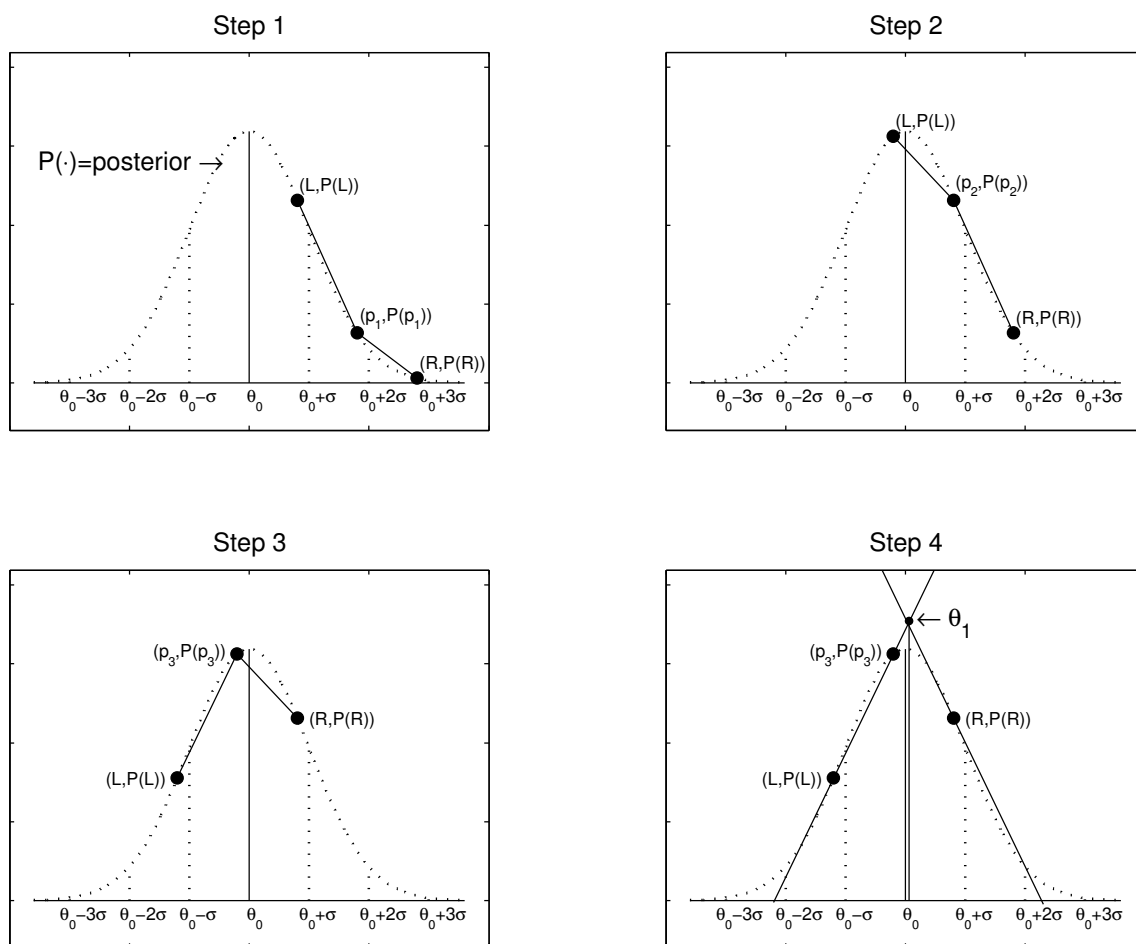


Figure 5. Use three points to find a mode (mean). The dashed line is proportional to the target distribution of the direction, given all other parameters.

Thus, given $x \in [a, b]$, we can calculate the range of k based on the supports corresponding to the functions $\phi_{j,k}(x)$ and $\psi_{j,k}(x)$ intersecting $[a, b]$,

$$[\lceil a2^j \rceil - 2N + 1, \lfloor b2^j \rfloor], \quad [\lceil a2^j \rceil - N, \lfloor b2^j \rfloor + N - 1], \quad (3.33)$$

where $\lfloor x \rfloor = \max\{n \in \mathbb{Z}; n \leq x\}$, and $\lceil x \rceil = \min\{n \in \mathbb{Z}; n \geq x\}$.

3.5.2 Initialization and hyperparameters settings

It is important to initialize the direction parameter θ since doing so affects the initialization of the coefficients $c_{0,k}$ and $w_{j,k}$. To obtain a good initial direction, we implemented preliminary MCMC runs with 1000 iterations for several starting directions, where we choose a direction from these directions by minimizing the residual sums of squares calculated based on the estimated mean function corresponding to each starting direction. Motivated by the orthogonality of wavelet functions we propose estimators of the coefficients that are based on a "quadrature" type estimate as follows (Hart, 1997):

$$\begin{aligned} \hat{c}_{0,k} &= \sum_{i=1}^n y_i (s_i - s_{i-1}) \phi_{0,k} \left(\frac{s_i + s_{i-1}}{2} \right) \\ &\approx \sum_{i=1}^n y_i \int_{s_{i-1}}^{s_i} \phi_{0,k}(z) dz, \end{aligned} \quad (3.34)$$

and

$$\begin{aligned} \hat{w}_{j,k} &= \sum_{i=1}^n y_i (s_i - s_{i-1}) \psi_{j,k} \left(\frac{s_i + s_{i-1}}{2} \right) \\ &\approx \sum_{i=1}^n y_i \int_{s_{i-1}}^{s_i} \psi_{j,k}(z) dz, \end{aligned} \quad (3.35)$$

where $z_{(i)}$ is the i th smallest of the ordered $X\beta$, $s_0 = z_{(1)}$, $s_i = \frac{z_{(i)} + z_{(i+1)}}{2}$, $i = 1, \dots, n-1$, $s_n = z_{(n)}$, and $\frac{s_i + s_{i-1}}{2} = \frac{z_{(i-1)} + 2z_{(i)} + z_{(i+1)}}{2}$ for $z_{(1)} < \dots < z_{(n)}$. The initial values of τ and α are generated from the priors. An initial choice of a variance of the normal proposal distribution for the direction parameter θ is

$$\hat{\sigma}_\theta^2 = \frac{\zeta}{n-1} \sum_{i=1}^n (y_i - \hat{r}_i)^2, \quad (3.36)$$

where ζ is arbitrary and \hat{r}_i is the wavelet series approximation corresponding to a starting direction. With the initial direction given, the variance $\hat{\sigma}_0^2$ of the MCMC scheme is chosen doing different sets of initial MCMC runs until a value of $\hat{\sigma}_0^2$ is found that yields an acceptance rate around 60% for the Metropolis steps and over 70% for the independence sampler.

The parameters a_ν and a_τ are fixed at $a_\nu = a_\tau = 1/2$ and the parameters b_ν and b_τ are fixed at $b_\nu = b_\tau = 1$ so that the priors of σ and $\sqrt{\tau}$ have infinite means (Hart, 1997). The parameters a_α and b_α are fixed at $a_\alpha = b_\alpha = 1$, meaning that α has a $U(0, 1)$ prior distribution.

CHAPTER IV

SIMULATION STUDIES AND AN APPLICATION

To calculate the ranges of the translation parameter k , we need the support of the indices for the true direction. Since the true direction θ is unknown, we use a maximum support of indices $Z = \mathbf{X}T(\theta)$ (A-8 in Appendix A) or an adjusted support, $[z_{(1)}, z_{(n)}]$, obtained from the indices of a direction updated by MCMC schemes.

We perform simulation studies on two non-linear functions, a cosine function which is represented by only the smooth parts of the wavelet series, and a Doppler function which is represented by the full wavelet representations. We compare performances of the Metropolis algorithm and the independent samplers for the direction parameter θ .

In these simulation studies we repeat the MCMC schemes twice with the first $T_1 = 1000$ iterations to initialize the direction parameter θ and the second $T_2 = 10000$ iterations including a burn-in period (1000 or 2000). We simulated 20 datasets for each of the two functions.

4.1 Simulation study for fixed resolution

The cosine function is represented only by smooth parts in the wavelet series (3.1) and the Doppler function is represented by the wavelet representations (3.1 or 3.2) with $m_0 = 5$ or 6. In the cosine function the supports of the indices for all directions are very close to that of the true direction. In the Doppler function we propose the maximum support and the adjusted supports.

For notational simplicity, suppose that Z denotes $\mathbf{X}\beta$ where $\mathbf{X} = (X_1, \dots, X_p)$ is $n \times p$, $\beta = (\beta_1, \dots, \beta_p)^T$ is $p \times 1$. The simulation data (z_i, y_i) , $i = 1, \dots, n$, are independent, where $y_i = r(z_i) + \varepsilon_i$, $z_i = X^i\beta = x_{i1}\beta_1 + \dots + x_{ip}\beta_p$, and $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$.

Example 1. The number of observations is $n = 200$ and the direction parameter is 2×1 . The function is $r(z_i) = \cos(z_i)$ and each covariate is independently generated from a normal distribution with zero mean. Table 1 shows the results for the function, the error variance, and the direction, with three different directions and three different error variances. Figures 6-11 show boxplots for the posterior means of the direction for the Metropolis algorithm and two independence samplers over the 20 repetitions, and the estimates of the mean function $\cos(z_i)$. The posterior distributions of the wavelet coefficients $c_{0,k}$ are shown in Figures 12-14 for one of the 20 repetitions.

Example 2. The number of observations is $n = 200$ and the direction parameter is 2×1 . The function is $r(z_i) = 2\sqrt{z_i(1-z_i)} \sin[(2.1\pi/(z_i + 0.05))]$, $i = 1, \dots, n$ and $n = 200$, for $0 < z_i < 1$, and each covariate is independently generated from a uniform distribution on $[0, 0.45]$. Tables 2-3 show the results with one direction and three different error variances for the Metropolis and two independence samplers with the maximum resolutions $m_0 = 5$ or 6 fixed. Figures 15-38 show the posterior means of the direction for the Metropolis algorithm and two independence samplers for the 20 repetitions, the estimated mean Doppler function, and the posterior distributions of the wavelet coefficients $w_{4,k}$ for one of the 20 repetitions.

In all tables and all figures, for convenience, let θ_P denote a previous value, θ_L denote a mode calculated based on the two lines in Figure 5, and θ_M denote a mode calculated based on the Matlab command *fminsearch*.

Table 1. Simulation results for $y_i = \cos(z_i) + \varepsilon_i$ where $z_i = X^i T(\theta)$. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined based on an acceptance rate (see Subsection 3.5.2).

Metropolis algorithm, $N(\theta_P, \hat{\sigma}_1^2)$

(θ, σ) true value	θ		σ		r	
	bias	mse	bias	mse	Ibias	Imse
(0.35, 0.02)	2.12e-4	3.31e-6	8.52e-2	7.26e-3	3.63e-4	5.43e-5
(0.35, 0.5)	3.00e-3	2.29e-3	1.10e-2	8.12e-4	1.36e-2	1.26e-2
(0.35, 1)	-6.24e-3	7.60e-3	2.37e-2	2.68e-3	-2.52e-3	5.79e-2
(2.54, 0.02)	4.26e-4	1.80e-6	8.52e-2	7.26e-3	3.40e-4	8.98e-5
(2.54, 0.5)	1.47e-2	1.39e-3	1.08e-2	7.84e-4	8.29e-3	1.43e-3
(2.54, 1)	1.78e-2	6.68e-3	1.87e-2	2.45e-3	-2.09e-2	5.54e-2
(4.72, 0.02)	-2.34e-4	3.64e-6	8.50e-2	7.22e-3	3.30e-4	5.06e-5
(4.72, 0.5)	-1.11e-2	1.66e-3	9.33e-3	8.21e-4	1.24e-2	1.37e-2
(4.72, 1)	8.76e-4	7.91e-3	2.76e-2	2.92e-3	-9.40e-4	6.01e-2

Independence sampler, $N(\theta_L, \hat{\sigma}_2^2)$

(0.35, 0.02)	1.98e-4	1.68e-6	8.53e-2	7.28e-3	-2.10e-5	3.48e-5
(0.35, 0.5)	4.55e-3	2.28e-3	7.06e-3	6.91e-4	9.00e-3	1.09e-2
(0.35, 1)	-1.24e-2	6.15e-3	9.51e-3	2.09e-3	-2.58e-3	5.78e-2
(2.54, 0.02)	7.46e-5	2.12e-6	8.51e-2	7.24e-3	-3.51e-5	3.89e-5
(2.54, 0.5)	1.21e-2	1.21e-3	6.13e-3	6.86e-4	8.40e-3	1.16e-6
(2.54, 1)	1.32e-2	7.26e-3	6.45e-3	2.00e-3	6.19e-3	4.77e-2
(4.72, 0.42)	-1.69e-4	2.77e-6	8.50e-2	7.23e-3	-2.86e-5	3.09e-5
(4.72, 0.5)	-1.24e-2	1.71e-3	5.64e-3	7.31e-4	8.60e-3	1.19e-2
(4.72, 1)	-5.66e-3	6.93e-3	1.17e-2	2.16e-3	1.17e-2	4.32e-2

Independence sampler, $N(\theta_M, \hat{\sigma}_2^2)$

(0.35, 0.02)	1.84e-4	1.70e-6	8.53e-2	7.28e-3	-2.09e-5	3.48e-5
(0.33, 0.5)	4.67e-3	2.29e-3	7.14e-3	6.94e-4	9.01e-3	1.09e-2
(0.35, 1)	-1.27e-2	6.13e-3	9.48e-3	2.09e-2	1.03e-2	4.51e-2
(2.54, 2.02)	6.47e-5	2.11e-6	8.51e-2	7.24e-3	-3.51e-5	3.89e-5
(2.54, 0.5)	1.21e-2	1.21e-3	6.14e-3	6.86e-4	8.40e-3	1.16e-2
(2.54, 1)	1.34e-2	7.24e-3	6.51e-3	2.02e-3	6.10e-3	4.77e-2
(4.72, 0.02)	-1.61e-4	2.79e-6	8.51e-2	7.23e-3	-1.57e-5	3.10e-5
(4.72, 0.5)	-1.24e-2	1.71e-3	5.64e-3	7.31e-4	8.63e-3	1.19e-2
(4.72, 1)	1.63e-1	4.58e-1	1.18e-2	2.18e-3	1.20e-2	4.34e-2

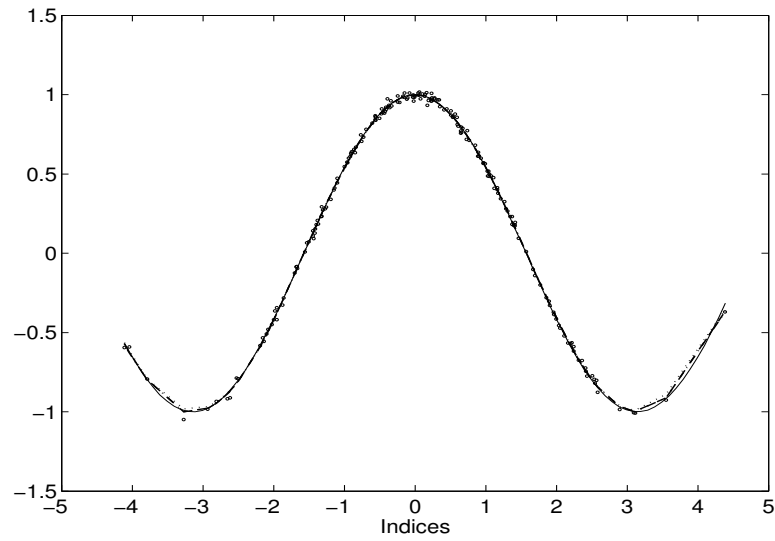


Figure 6. Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command fminsearch (dash-dot line), together with the true curve(solid line) on the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.02^2)$.

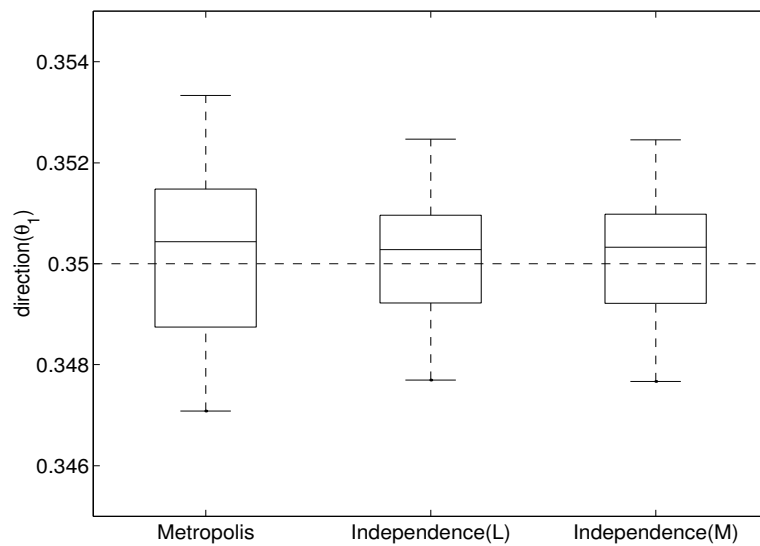


Figure 7. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.02^2)$. The box has lines at the lower quartile, median, and upper quartile values.

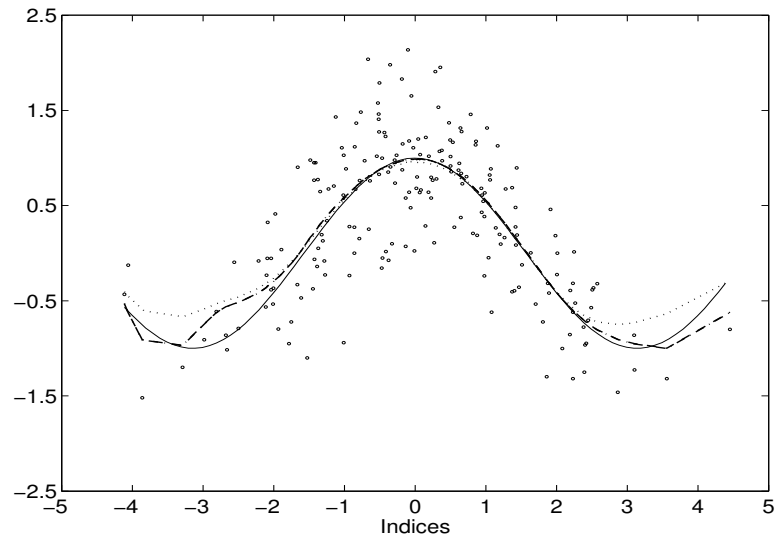


Figure 8. Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command `fminsearch` (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.5^2)$.

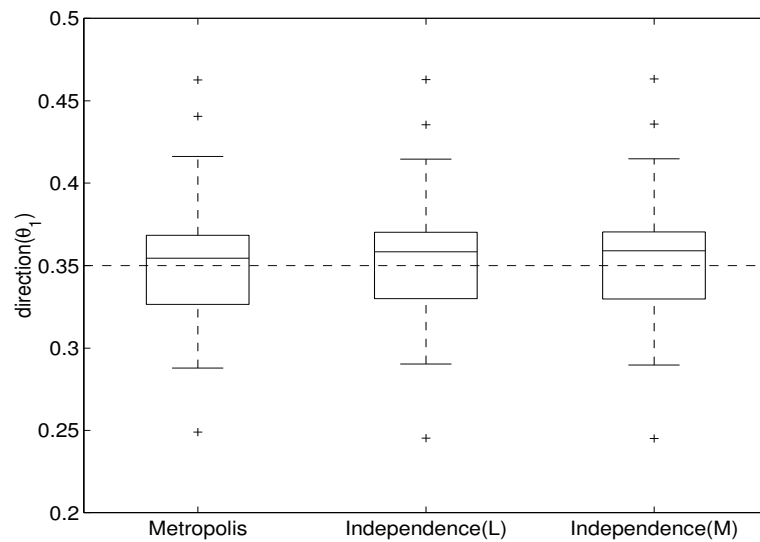


Figure 9. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.5^2)$. The box has lines at the lower quartile, median, and upper quartile values.

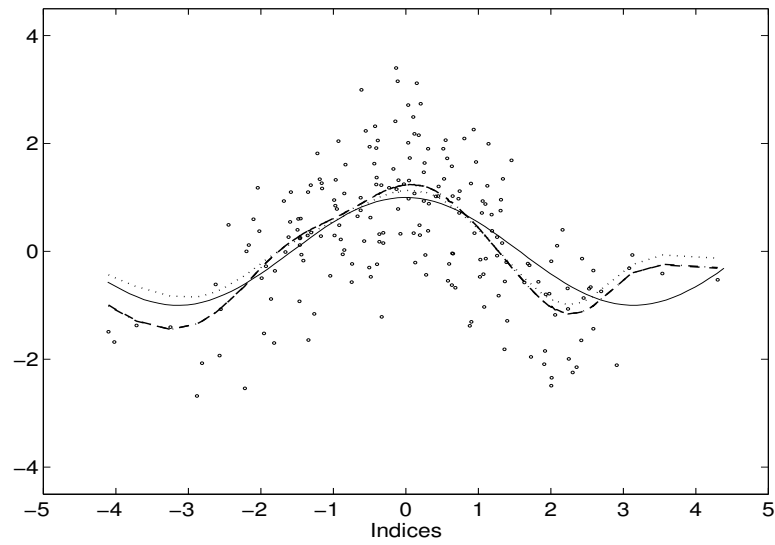


Figure 10. Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command `fminsearch` (dash-dot line) together with the true curve (solid line) on $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 1^2)$.

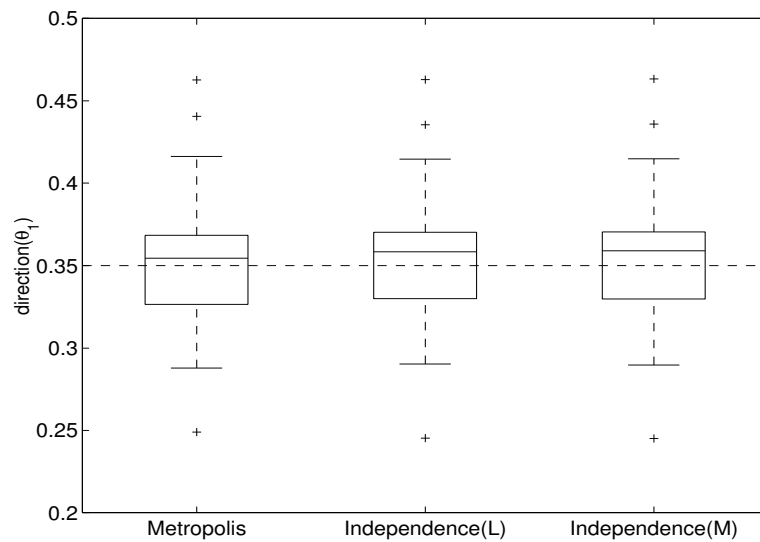


Figure 11. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 1^2)$. The box has lines at the lower quartile, median, and upper quartile values.

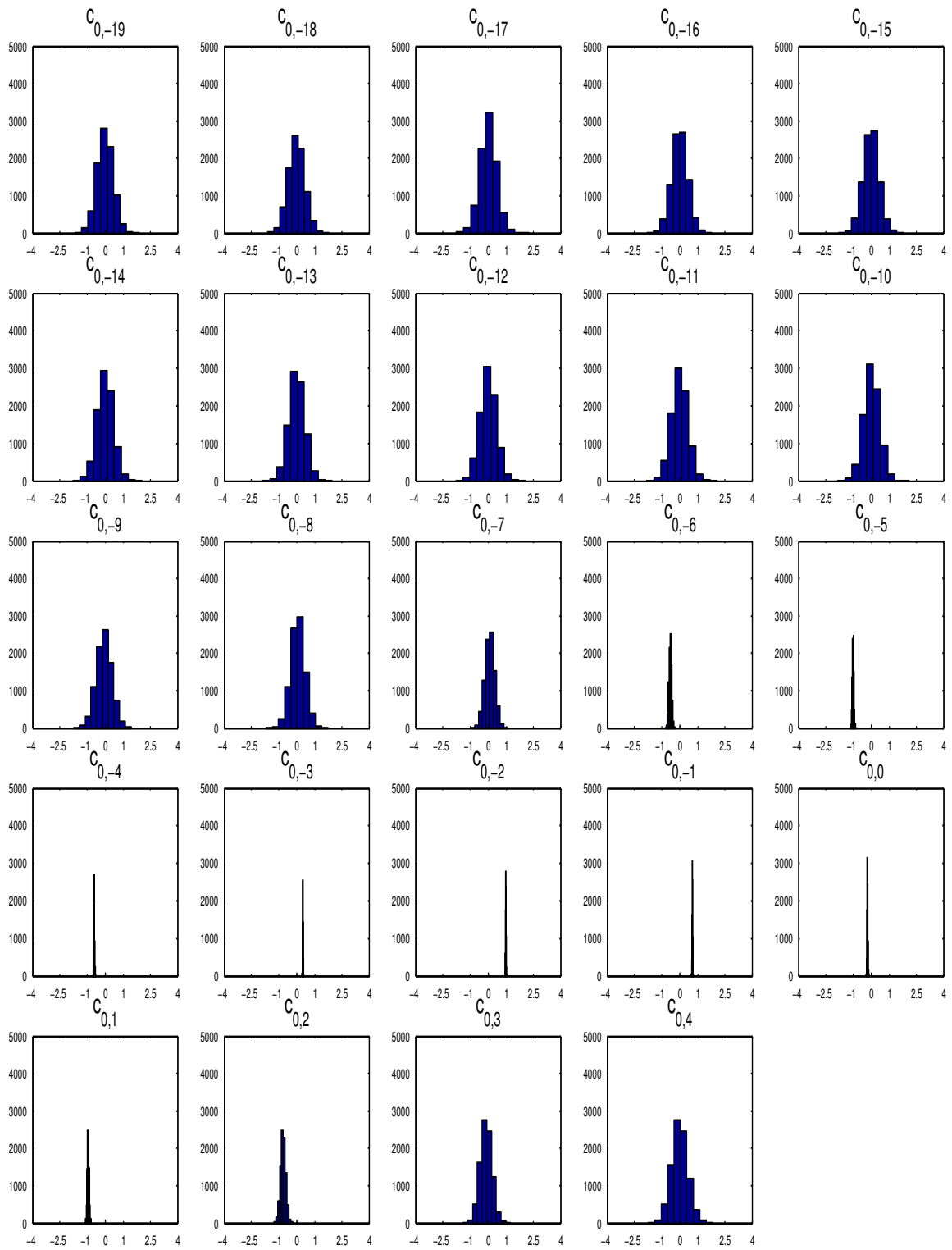


Figure 12. Histograms for the posterior distribution of $c_{0,k}$ for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \varepsilon_i$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.02^2)$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000.

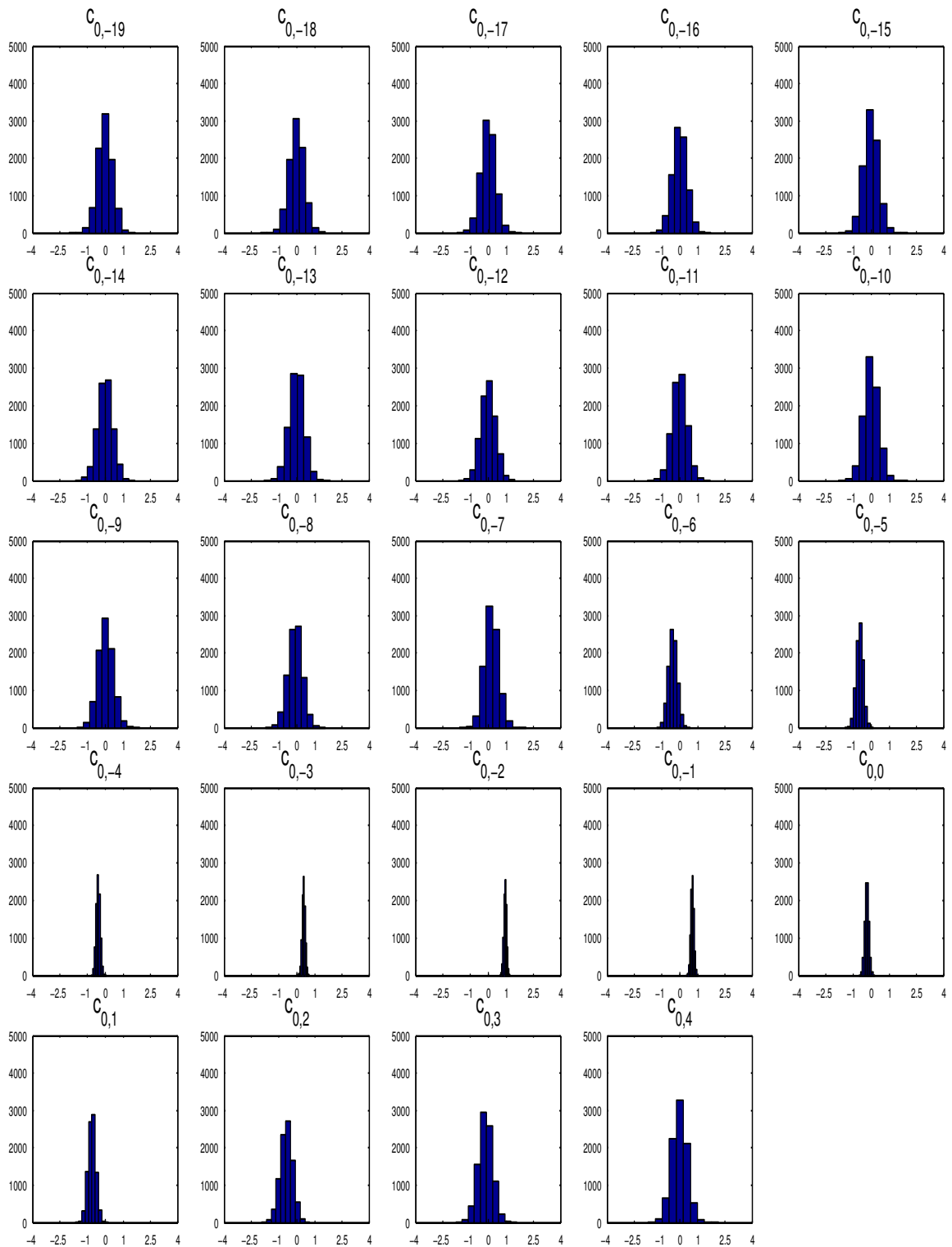


Figure 13. Histograms for the posterior distribution of $c_{0,k}$ for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \epsilon_i$, $\epsilon_i \stackrel{iid}{\sim} N(0, 0.5^2)$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000.

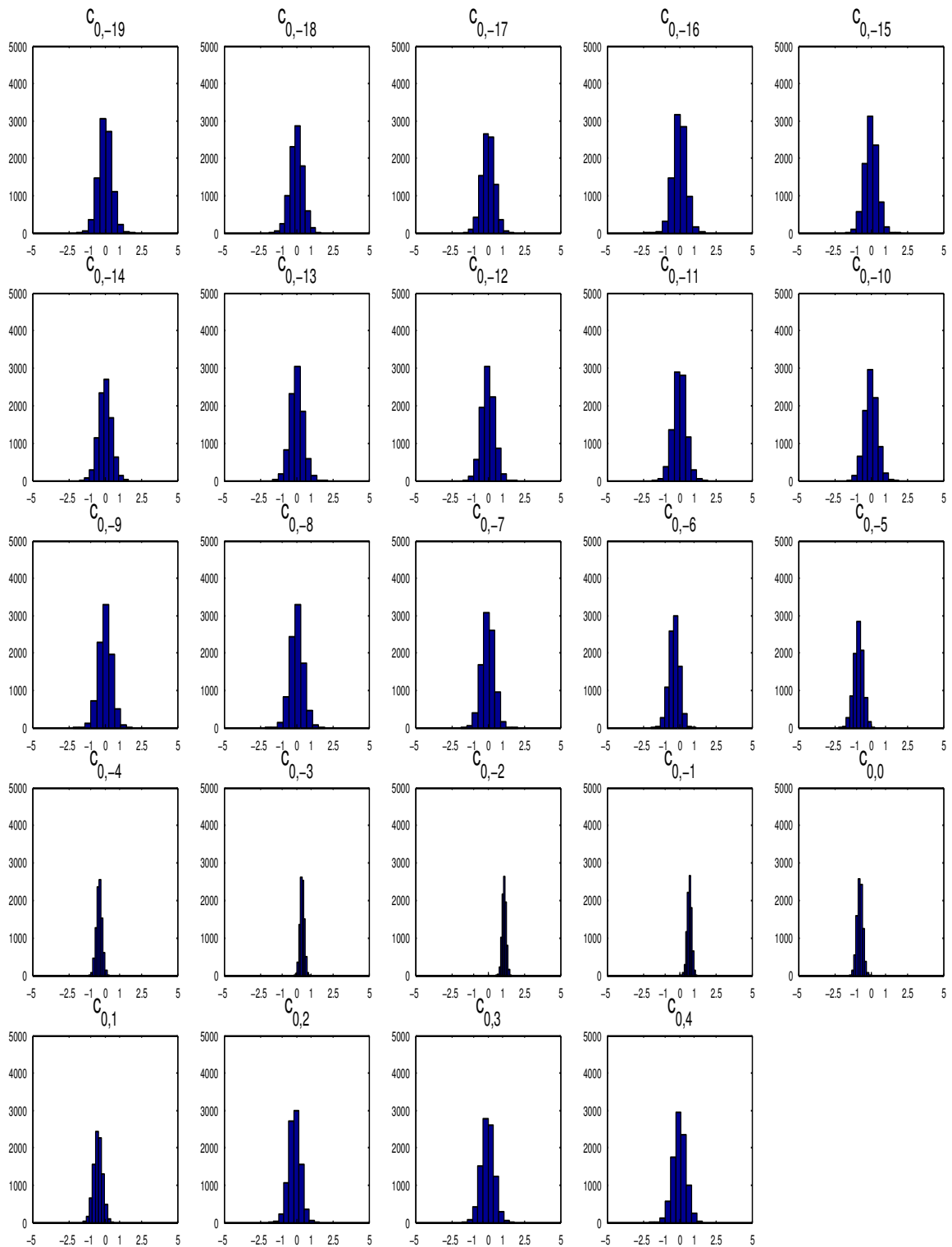


Figure 14. Histograms for the posterior distribution of $c_{0,k}$ for the true direction $\theta = 0.35$ and $y_i = \cos(z_i) + \epsilon_i$, $\epsilon_i \stackrel{iid}{\sim} N(0, 1^2)$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000.

Table 2. Simulation results for $y_i = 2\sqrt{z_i(1-z_i)} \sin\left(\frac{2.1\pi}{z_i+0.05}\right) + \varepsilon_i$ with the maximum support and with no shrinkage where $z_i = X^i T(\theta)$. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined based on an acceptance rate (see Subsection 3.5.2).

Metropolis algorithm, $N(\theta_P, \sigma_1^2)$

resolution	(θ, σ) true value	θ		σ		r	
		bias	mse	bias	mse	Ibias	Imse
$m_0 = 5$	(0.35, 0.02)	-6.54e-3	4.31e-5	1.67e-1	2.80e-2	3.92e-4	1.80e-2
	(0.35, 0.5)	-5.84e-3	5.96e-5	3.16e-2	1.53e-3	8.10e-3	5.40e-2
	(0.35, 1)	-6.25e-4	2.63e-4	4.12e-3	2.42e-3	7.57e-3	1.45e-1
$m_0 = 6$	(0.35, 0.02)	1.06e-4	1.00e-7	1.06e-1	1.11e-2	-1.22e-4	2.64e-4
	(0.35, 0.5)	2.11e-3	2.26e-5	2.45e-3	5.14e-4	6.81e-3	6.55e-2

Independence sampler, $N(\theta_L, \sigma_2^2)$

$m_0 = 5$	(0.35, 0.02)	-6.45e-3	4.17e-5	1.68e-1	2.81e-2	3.91e-4	1.80e-2
	(0.35, 0.5)	-5.70e-3	5.84e-5	3.18e-2	1.55e-3	8.10e-3	5.42e-2
	(0.35, 1)	-4.94e-4	2.01e-4	3.64e-3	2.39e-3	7.41e-3	1.47e-1
$m_0 = 6$	(0.35, 0.02)	7.52e-5	7.72e-8	1.06e-1	1.11e-2	-1.22e-4	2.64e-4
	(0.35, 0.5)	2.32e-3	2.40e-5	2.69e-3	5.12e-4	6.74e-3	6.63e-2

Independence sampler, $N(\theta_M, \sigma_2^2)$

$m_0 = 5$	(0.35, 0.02)	-6.46e-3	4.19e-5	1.68e-1	2.81e-2	3.91e-4	1.80e-2
	(0.35, 0.5)	-5.71e-3	5.87e-5	3.18e-2	1.55e-3	8.09e-3	5.43e-2
	(0.35, 1)	-7.63e-4	2.04e-4	3.48e-3	2.40e-3	7.47e-3	1.47e-1

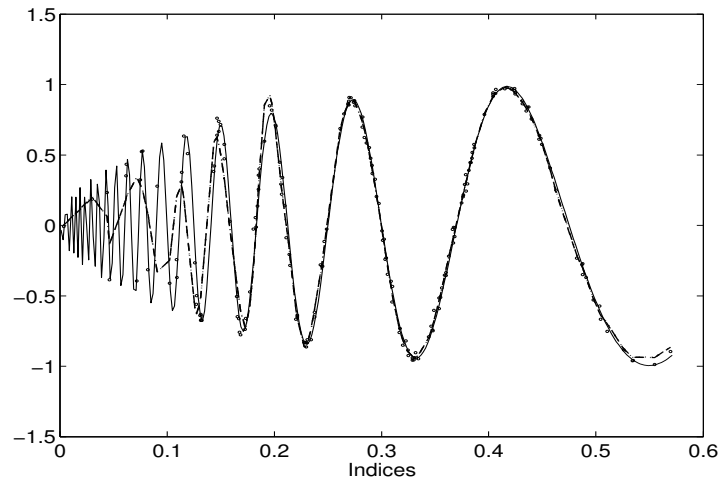


Figure 15. Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence based on two lines (dashed line), and the independence based on the Matlab command `fminsearch` (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$.

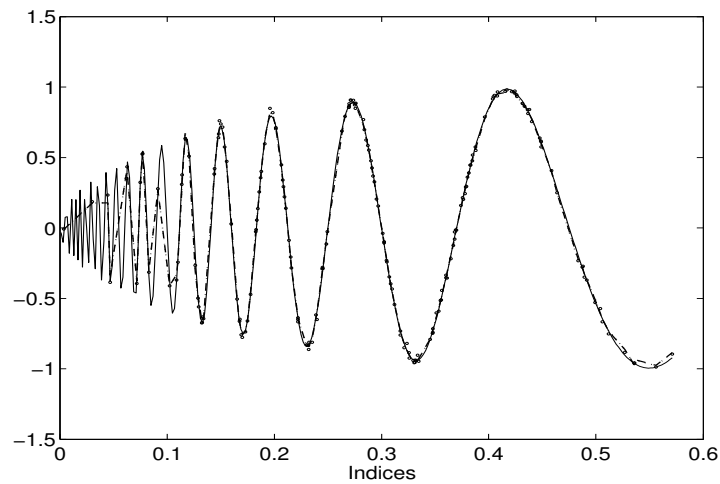


Figure 16. Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence based on two lines (dashed line), and the independence based on the Matlab command `fminsearch` (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 6$.

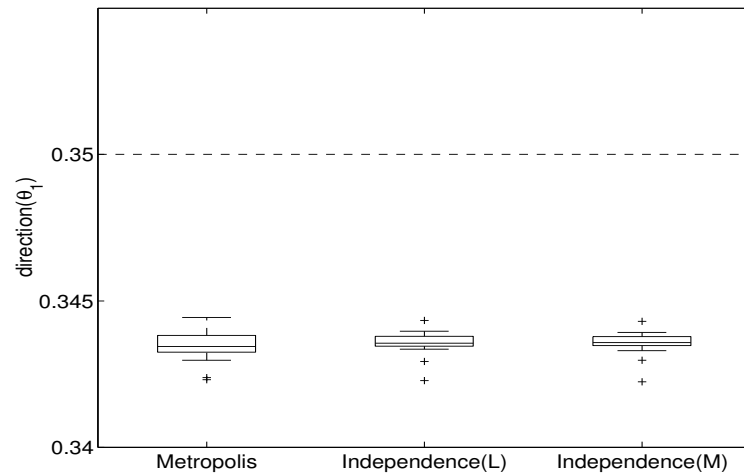


Figure 17. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

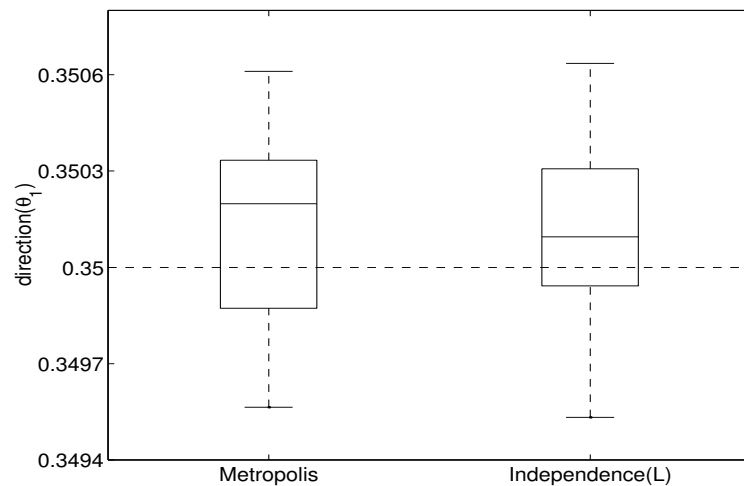


Figure 18. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 6$. The box has lines at the lower quartile, median, and upper quartile values.

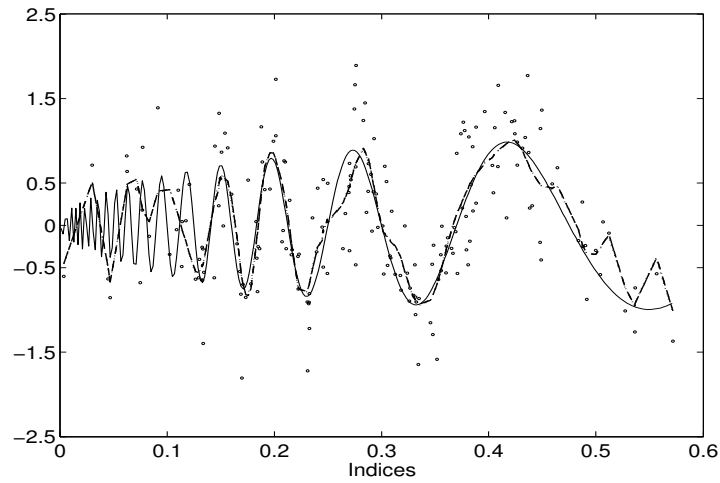


Figure 19. Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence based on two lines (dashed line), and the independence based on the Matlab command `fminsearch` (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$.

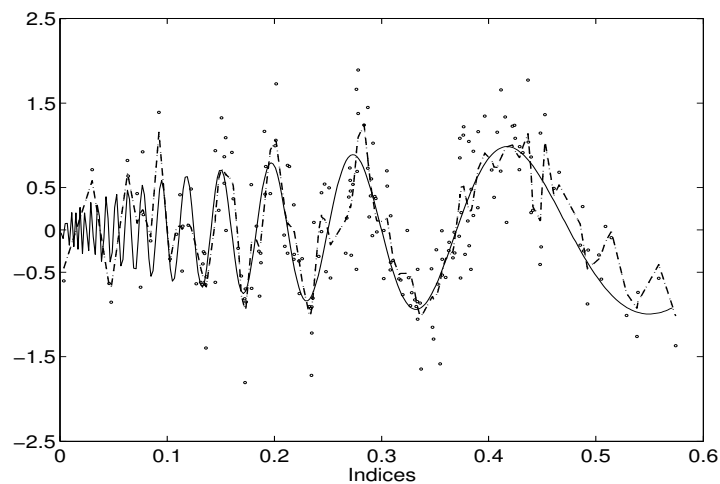


Figure 20. Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence based on two lines (dashed line), and the independence based on the Matlab command `fminsearch` (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 6$.

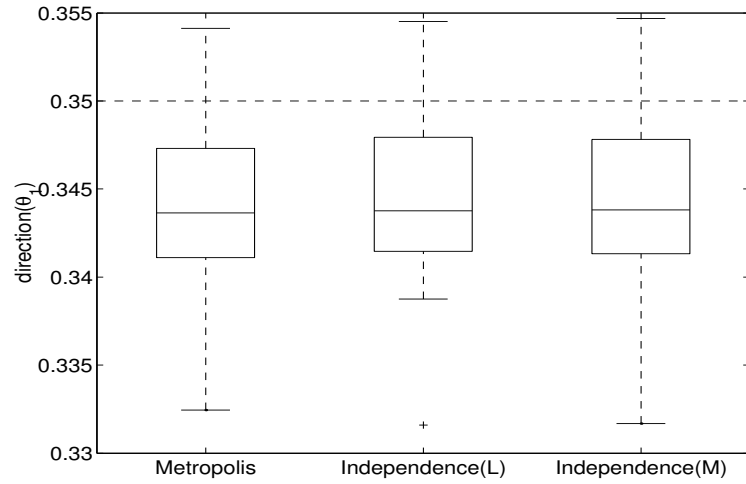


Figure 21. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

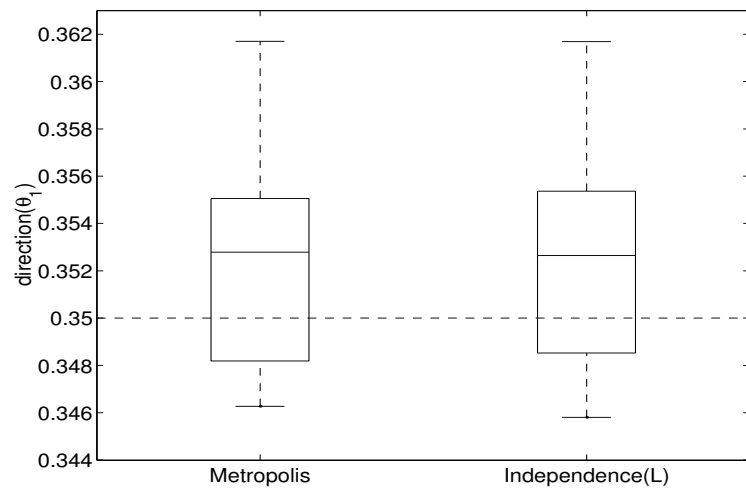


Figure 22. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 6$. The box has lines at the lower quartile, median, and upper quartile values.

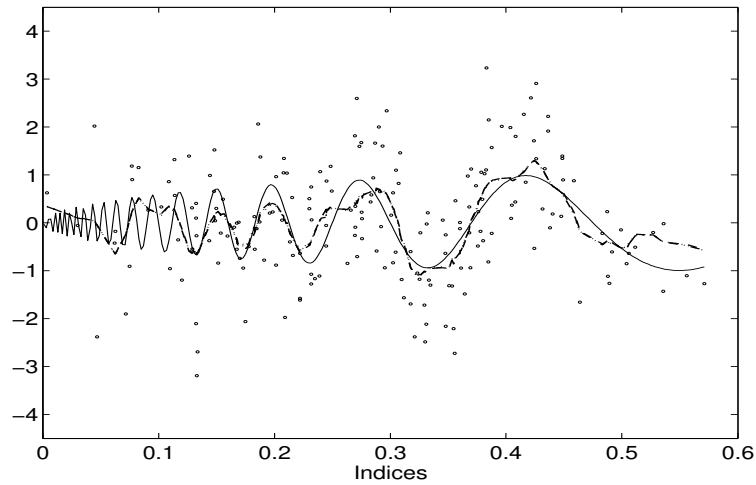


Figure 23. Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command `fminsearch` (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$.

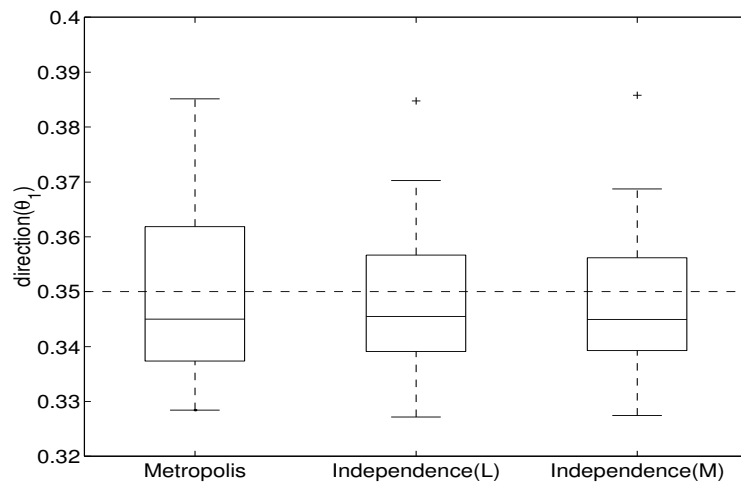


Figure 24. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

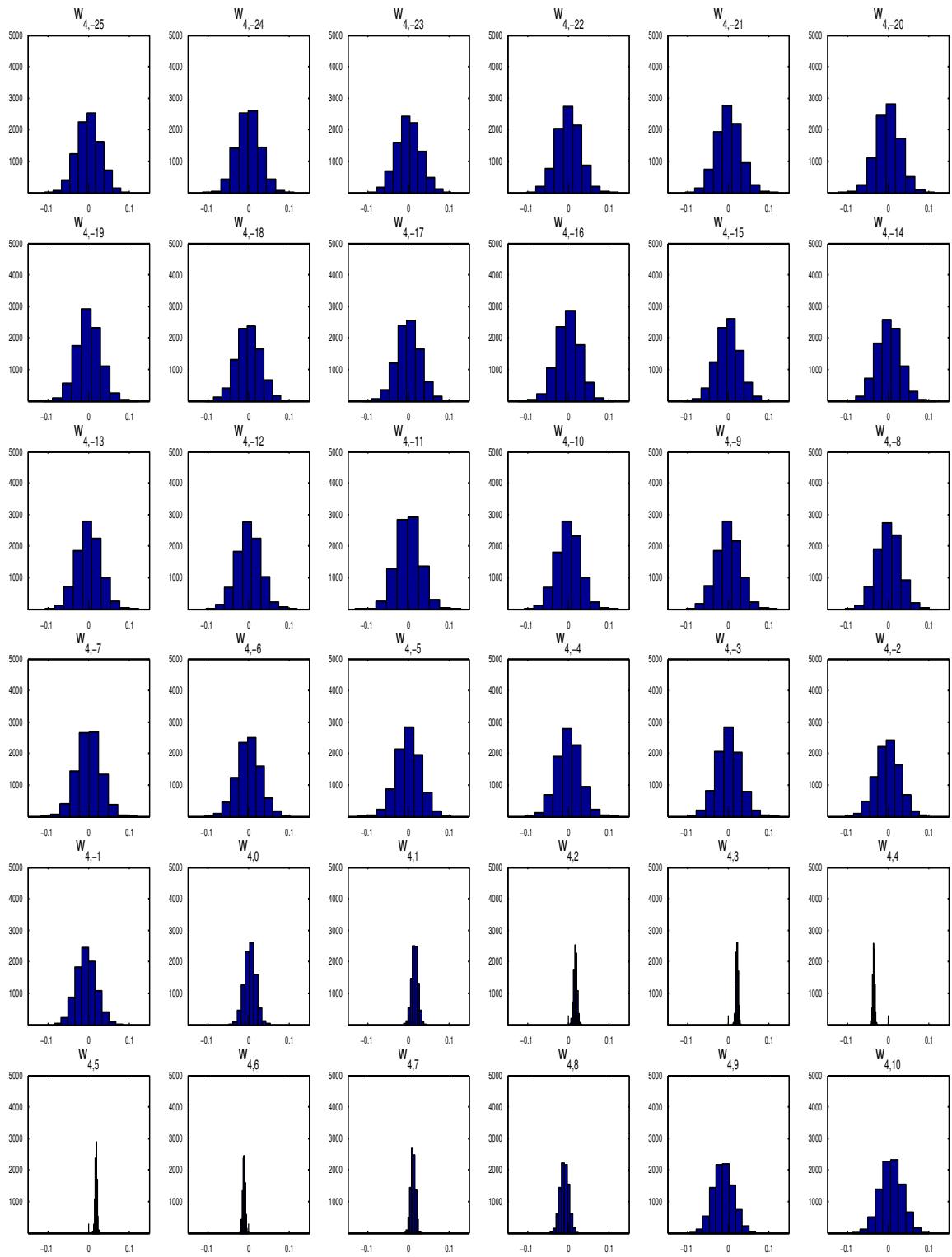


Figure 25. Histograms for the posterior distribution of $w_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000.

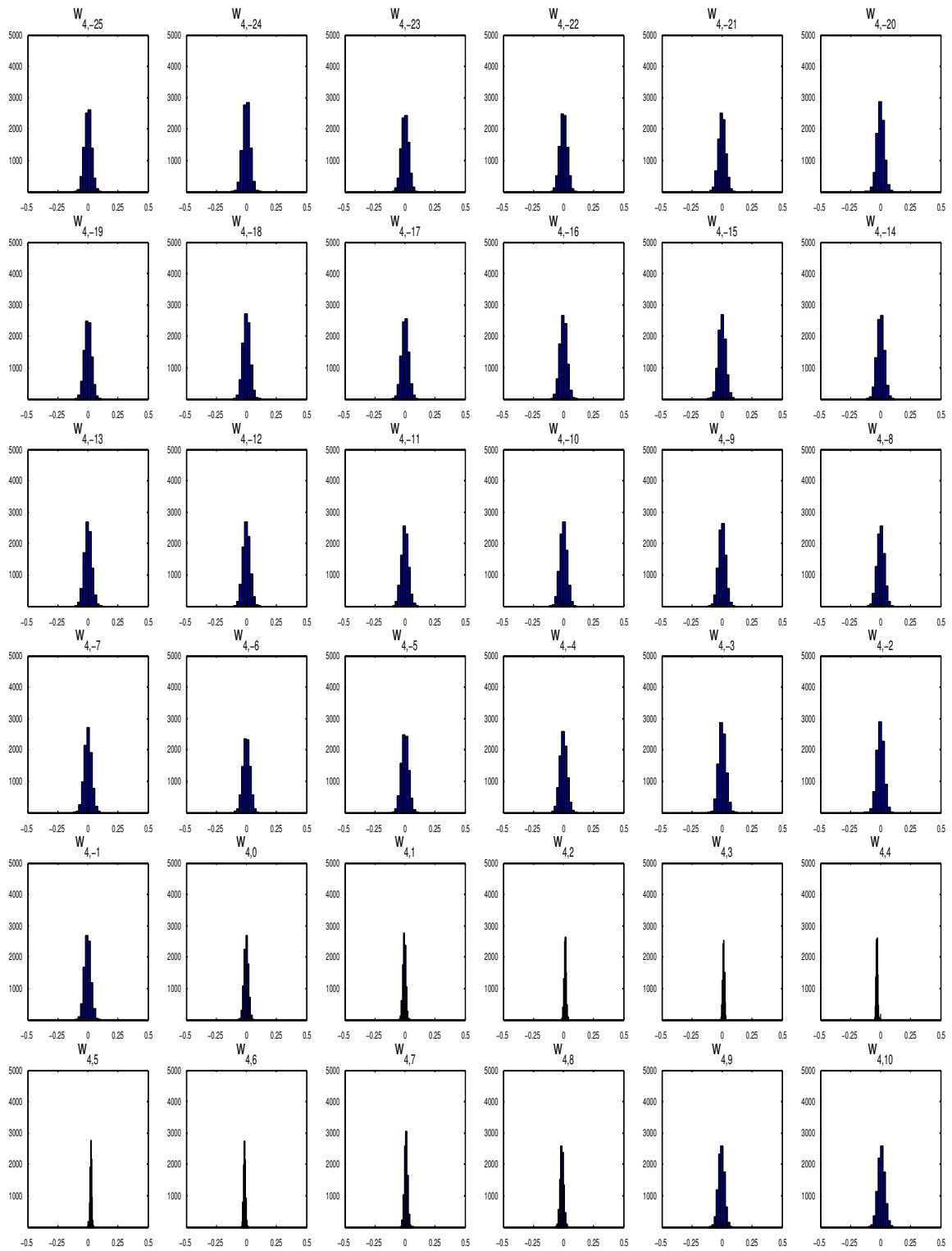


Figure 26. Histograms for the posterior distribution of $w_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000.

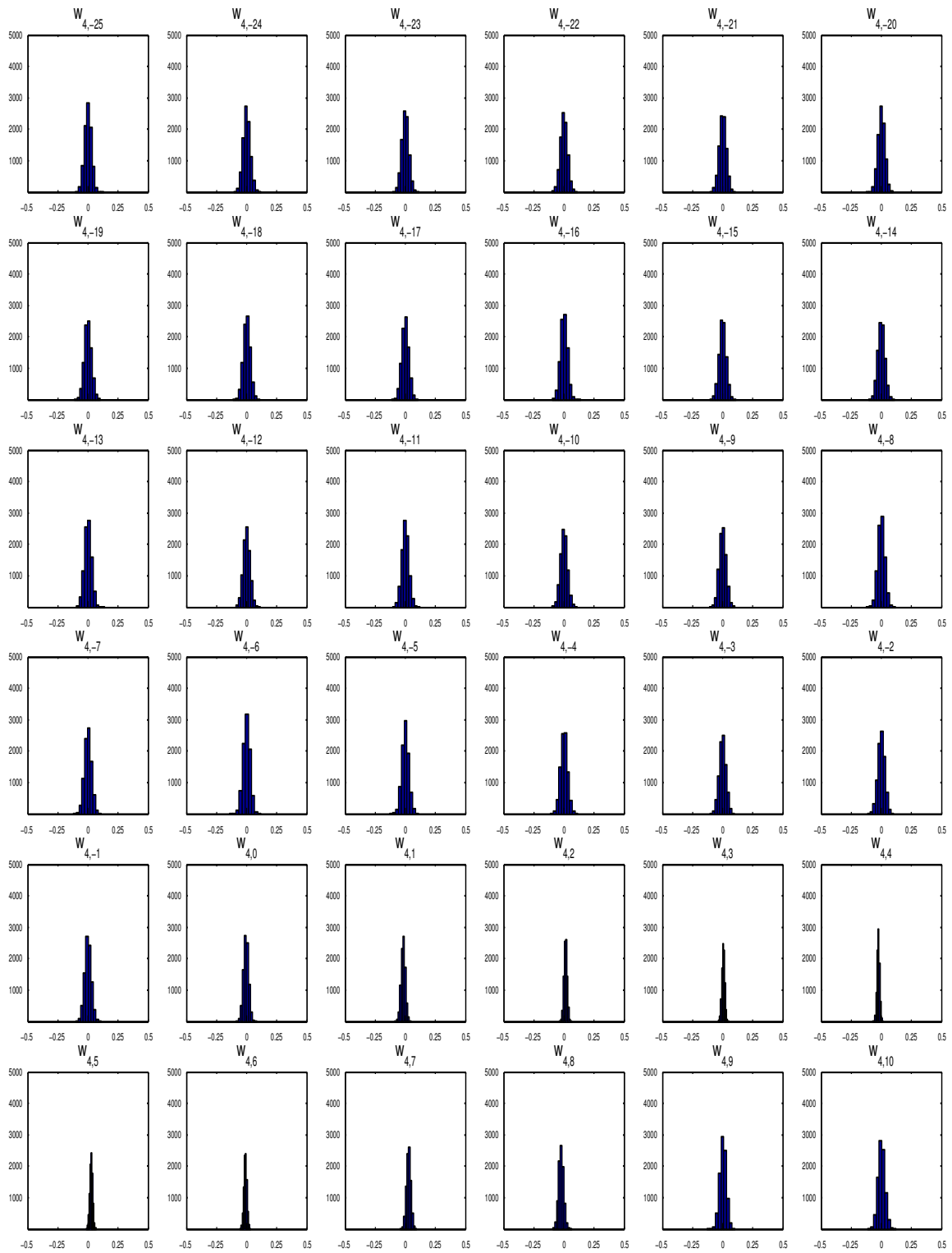


Figure 27. Histograms for the posterior distribution of $w_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the maximum support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000.

Table 3. Simulation results for $y_i = 2\sqrt{z_i(1-z_i)} \sin\left(\frac{2.1\pi}{z_i+0.05}\right) + \varepsilon_i$ with the adjusted support and with no shrinkage where $z_i = X^i T(\theta)$. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined based on an acceptance rate (see Subsection 3.5.2).

Metropolis algorithm, $N(\theta_P, \sigma_1^2)$

resolution	(θ, σ) true value	θ		σ		r	
		bias	mse	bias	mse	bias	mse
$m_0 = 5$	(0.35, 0.02)	-6.54e-3	4.31e-5	1.67e-1	2.80e-2	3.92e-4	1.80e-2
	(0.35, 0.5)	-5.84e-3	5.96e-5	3.16e-2	1.53e-3	8.10e-3	5.40e-2
	(0.35, 1)	-6.25e-4	2.63e-4	4.12e-3	2.42e-3	7.57e-3	1.45e-1
$m_0 = 6$	(0.35, 0.02)	1.47e-4	1.13e-7	1.06e-1	1.11e-2	-1.77e-4	2.44e-4

Independence sampler, $N(\theta_L, \sigma_2^2)$

$m_0 = 5$	(0.35, 0.02)	-6.45e-3	4.17e-5	1.68e-1	2.81e-2	3.91e-4	1.80e-2
	(0.35, 0.5)	-5.70e-3	5.84e-5	3.18e-2	1.55e-3	8.10e-3	5.42e-2
	(0.35, 1)	-4.94e-4	2.01e-4	3.64e-3	2.39e-3	7.41e-3	1.47e-1
$m_0 = 6$	(0.35, 0.02)	1.60e-4	8.74e-7	1.06e-1	1.11e-2	-2.03e-4	2.40e-4

Independence sampler, $N(\theta_M, \sigma_2^2)$

$m_0 = 5$	(0.35, 0.02)	-6.46e-3	4.19e-5	1.68e-1	2.81e-2	3.91e-4	1.80e-2
	(0.35, 0.5)	-5.71e-3	5.87e-5	3.18e-2	1.55e-3	8.09e-3	5.43e-2
	(0.35, 1)	-7.63e-4	2.04e-4	3.48e-3	2.40e-3	7.47e-3	1.47e-1

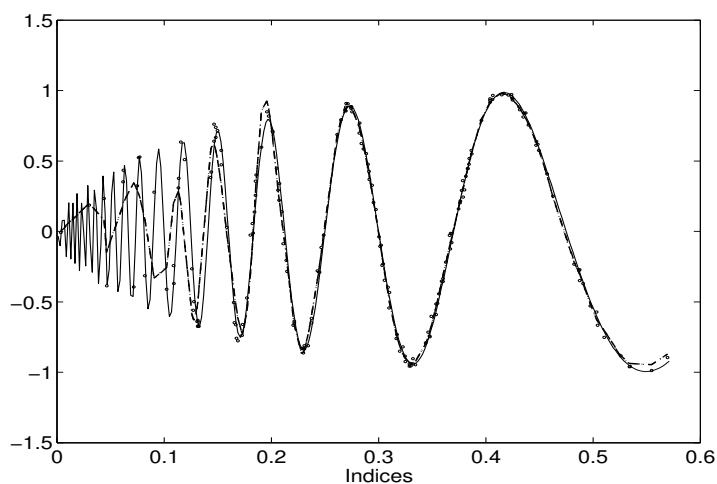


Figure 28. Posterior estimated mean functions for the Metropolis (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command `fminsearch` (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$.

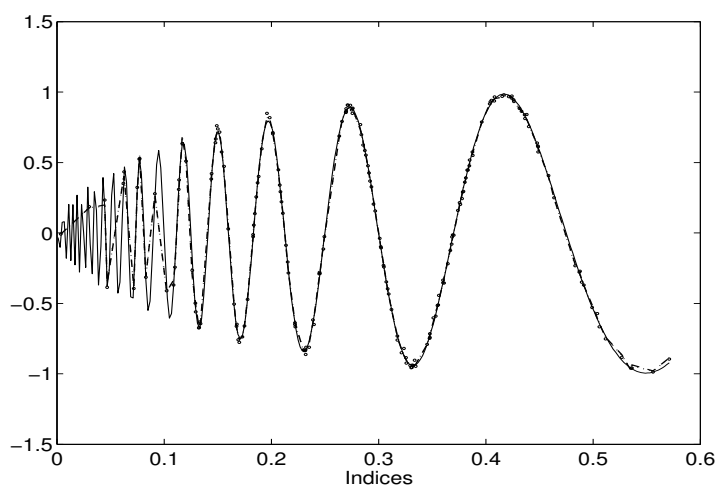


Figure 29. Posterior estimated mean functions for the Metropolis (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command `fminsearch` (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 6$.

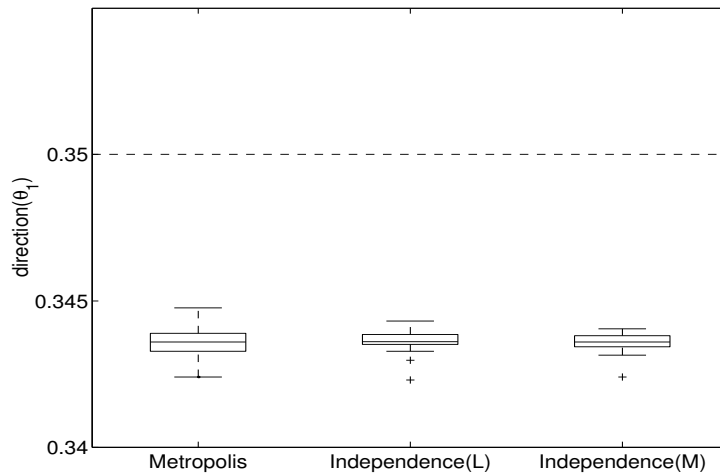


Figure 30. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

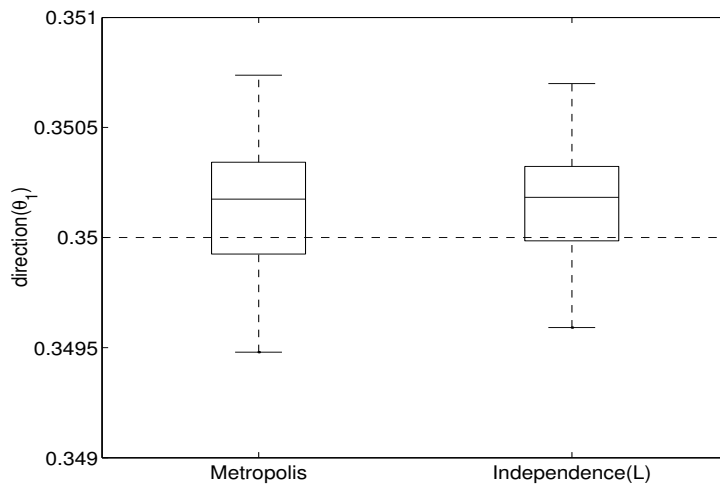


Figure 31. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 6$. The box has lines at the lower quartile, median, and upper quartile values.

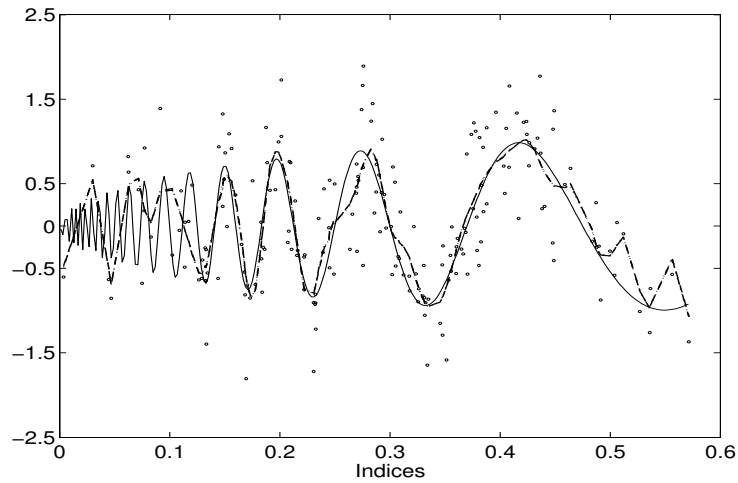


Figure 32. Posterior estimated mean functions for the Metropolis algorithm (dotted line), the independence sampler based on two lines (dashed line), and the independence sampler based on the Matlab command `fminsearch` (dash-dot line) together with the true curve (solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the support adjusted and no shrinkage and the maximum resolution $m_0 = 5$.

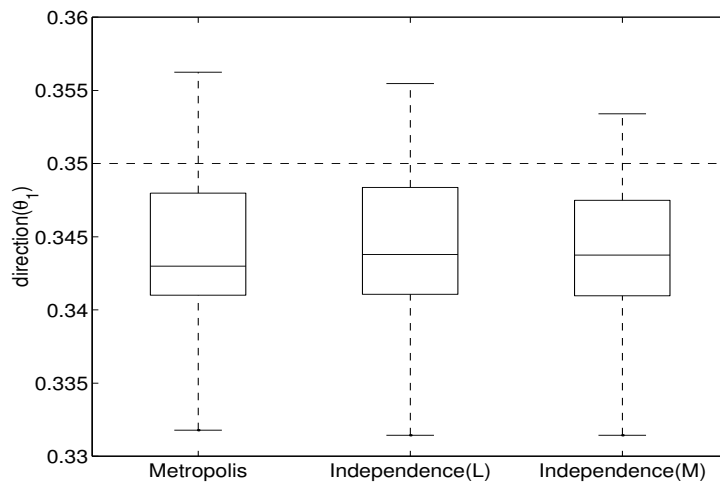


Figure 33. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the support adjusted and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

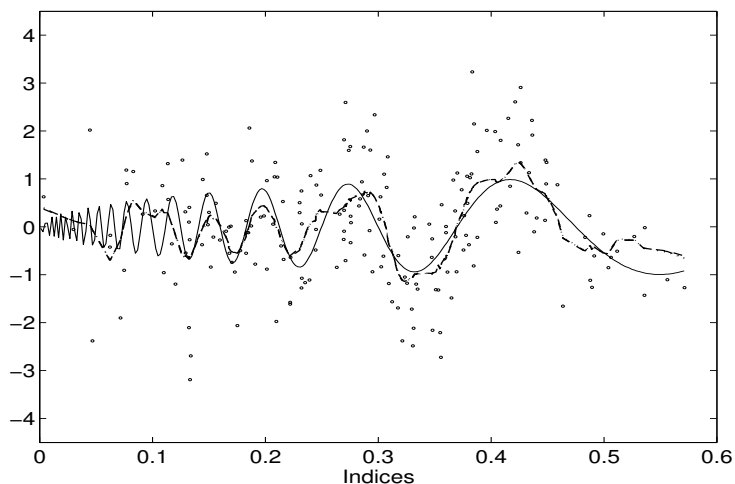


Figure 34. Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines(dashed line), and the independence sampler based on the Matlab command `fminsearch` (dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$.

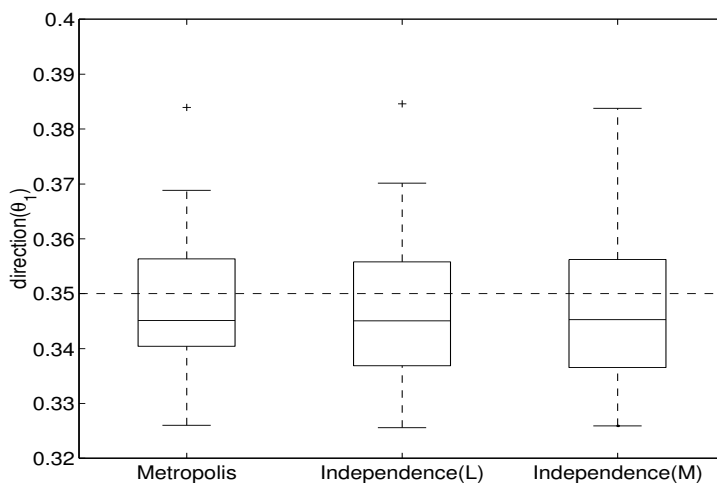


Figure 35. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

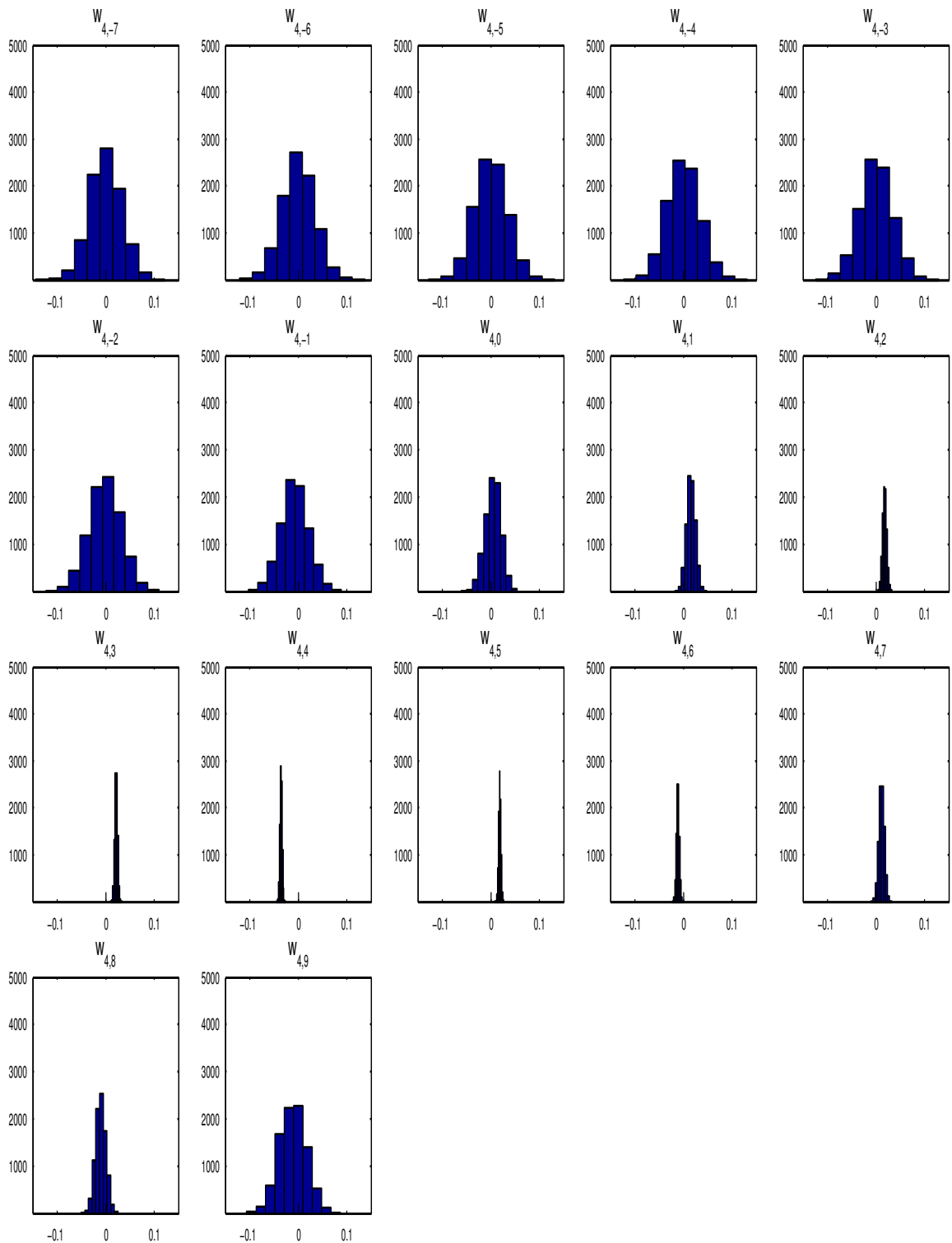


Figure 36. Histograms for the posterior distribution of $w_{4,k}$ for $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000.

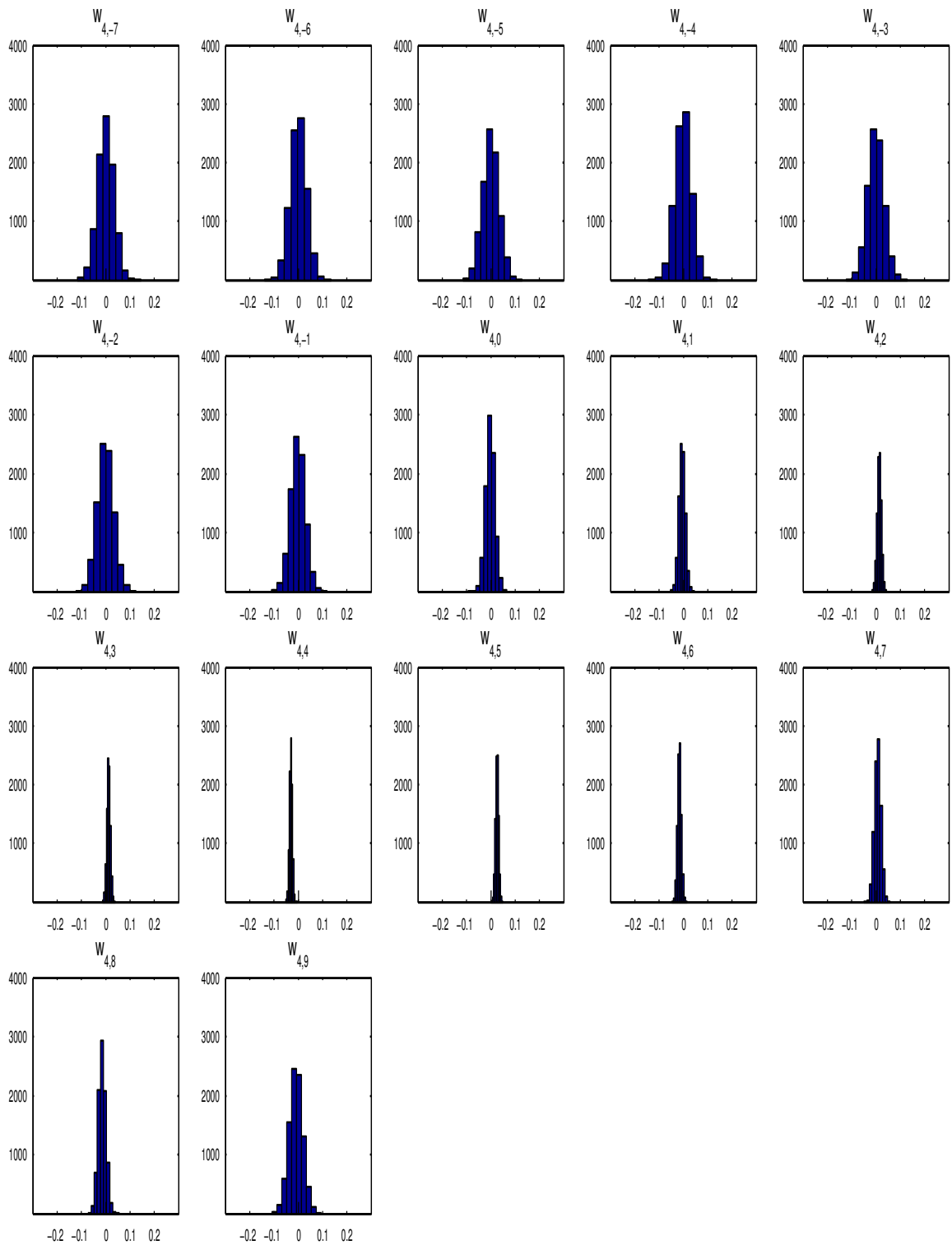


Figure 37. Histograms for the posterior distribution of $w_{4,k}$ for $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000.

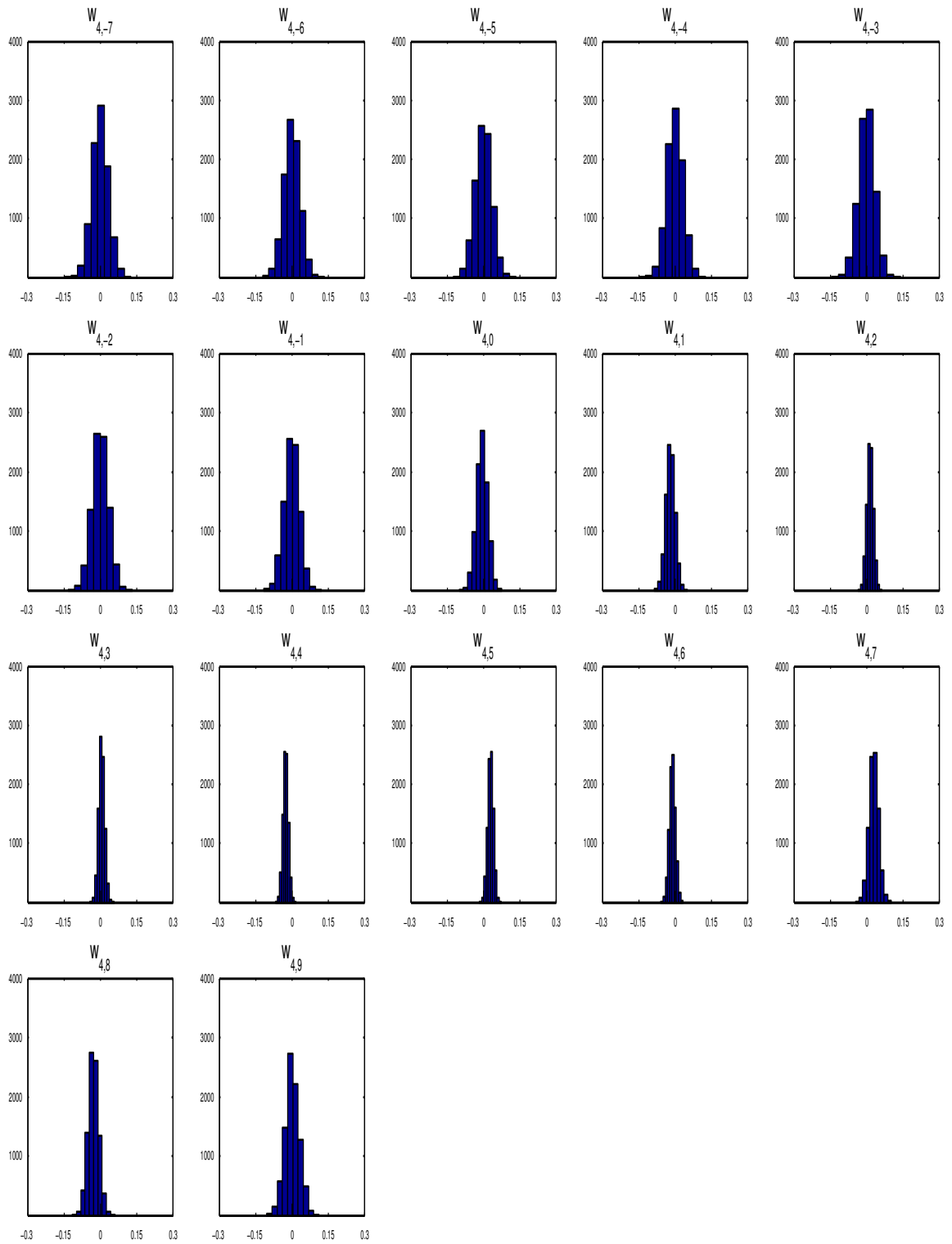


Figure 38. Histograms for the posterior distribution of $w_{4,k}$ for $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the adjusted support and no shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 9000.

4.2 Simulation study for wavelet shrinkage

To implement the parsimony of the wavelet representations, we can consider the Doppler function in Example 2 and represent it by the wavelet shrinkage series (3.5). Tables 4 and 5 show the results with one direction and three different error variances for the Metropolis and two independence samplers with the maximum resolutions $m_0 = 5$ or 6 fixed. Figures 39-60 show the posterior means of the direction for the Metropolis algorithm and two independence samplers for the 20 repetitions, the estimated mean Doppler function, and the posterior distributions of the indicators $s_{4,k}$ for one of the 20 repetitions.

Table 4. Simulation results for $y_i = 2\sqrt{z_i(1-z_i)} \sin\left(\frac{2.1\pi}{z_i+0.05}\right) + \varepsilon_i$ with the maximum support and with shrinkage where $z_i = X^i T(\theta)$. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined based on an acceptance rate (see Subsection 3.5.2).

Metropolis algorithm, $N(\theta_P, \sigma_1^2)$

resolution	(θ, σ) true value	θ		σ		r	
		bias	mse	bias	mse	bias	mse
$m_0 = 5$	(0.35, 0.02)	-6.32e-3	4.02e-5	1.71e-1	2.92e-2	4.64e-4	1.80e-2
	(0.35, 0.5)	-5.95e-3	6.17e-5	3.62e-2	1.87e-2	8.45e-3	5.36e-2
	(0.35, 1)	-3.93e-4	2.68e-4	1.42e-2	2.65e-3	7.94e-3	1.43e-1
$m_0 = 6$	(0.35, 0.02)	-1.76e-4	1.25e-7	1.46e-1	2.14e-2	-2.50e-4	2.84e-4

Independence sampler, $N(\theta_L, \sigma_2^2)$

$m_0 = 5$	(0.35, 0.02)	-6.51e-3	4.25e-5	1.70e-1	2.90e-2	7.94e-4	1.82e-2
	(0.35, 0.5)	-5.77e-3	6.13e-5	3.67e-2	1.91e-3	8.50e-3	5.39e-2
	(0.35, 1)	-4.09e-4	2.36e-4	1.28e-2	2.52e-3	7.75e-3	1.45e-1
$m_0 = 6$	(0.35, 0.02)	1.61e-4	1.77e-7	1.45e-1	2.11e-2	-2.20e-4	2.77e-4

Independence sampler, $N(\theta_M, \sigma_2^2)$

$m_0 = 5$	(0.35, 0.02)	-6.45e-3	4.18e-5	1.72e-1	2.94e-2	4.56e-4	1.80e-2
	(0.35, 0.5)	-5.68e-3	6.21e-5	3.67e-2	1.91e-3	8.43e-3	5.41e-2
	(0.35, 1)	5.96e-4	1.13e-3	1.81e-2	2.80e-3	5.97e-3	1.44e-1

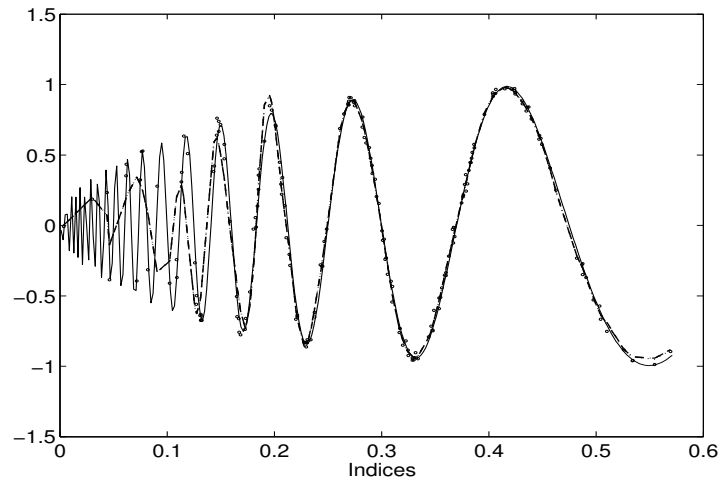


Figure 39. Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines(dashed line), and the independence sampler based on the Matlab command `fminsearch`(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$.

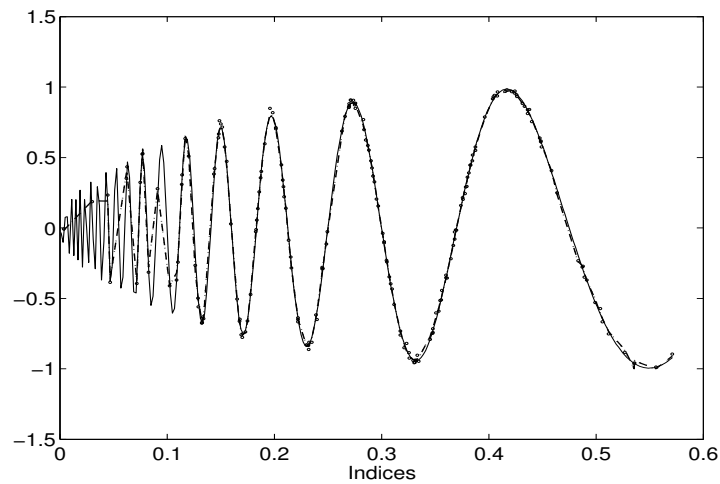


Figure 40. Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines(dashed line), and the independence sampler based on the Matlab command `fminsearch`(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 6$.

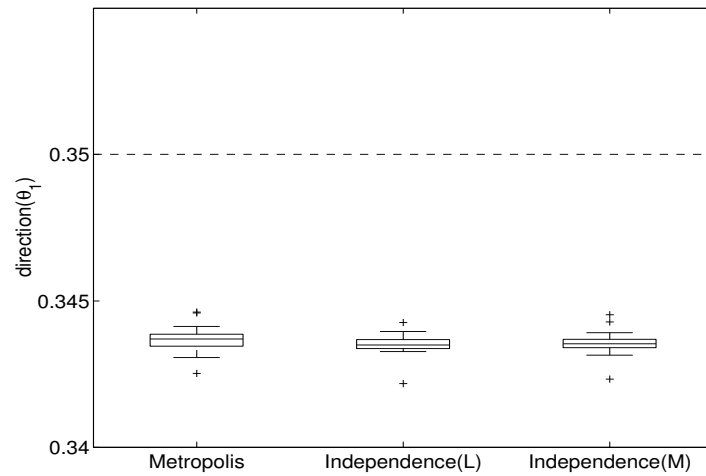


Figure 41. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

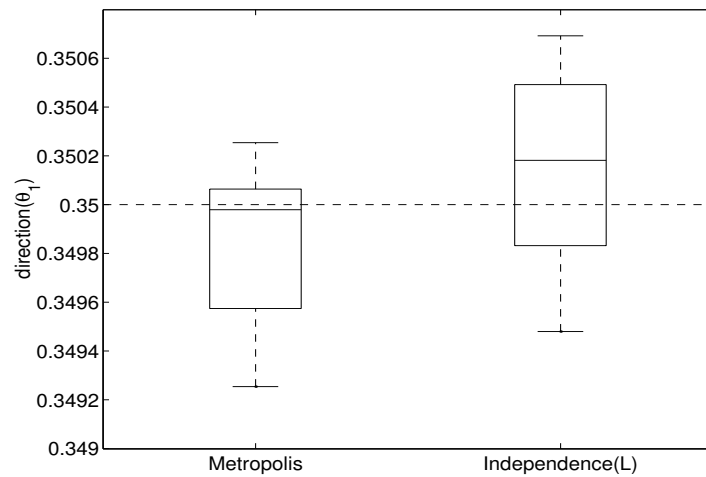


Figure 42. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 6$. The box has lines at the lower quartile, median, and upper quartile values.

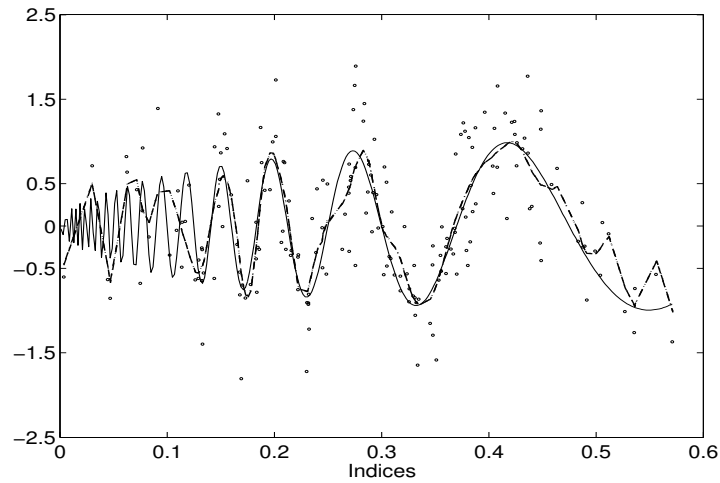


Figure 43. Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines(dashed line), and the independence sampler based on the Matlab command `fminsearch`(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$.

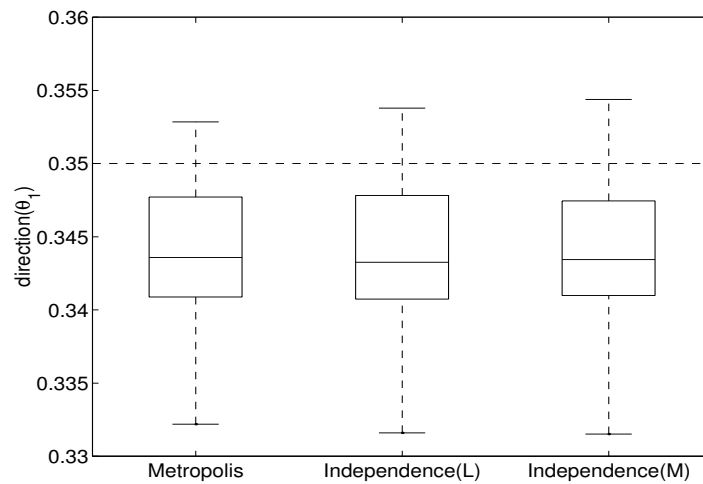


Figure 44. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

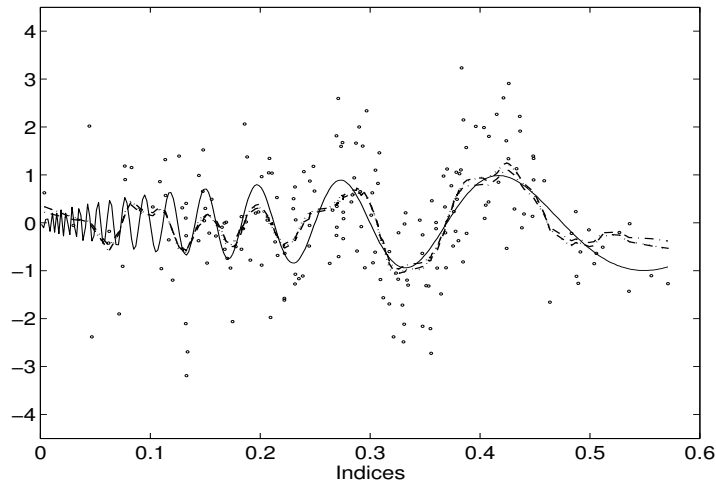


Figure 45. Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence sampler based on two lines(dashed line), and the independence sampler based on the Matlab command `fminsearch`(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$.

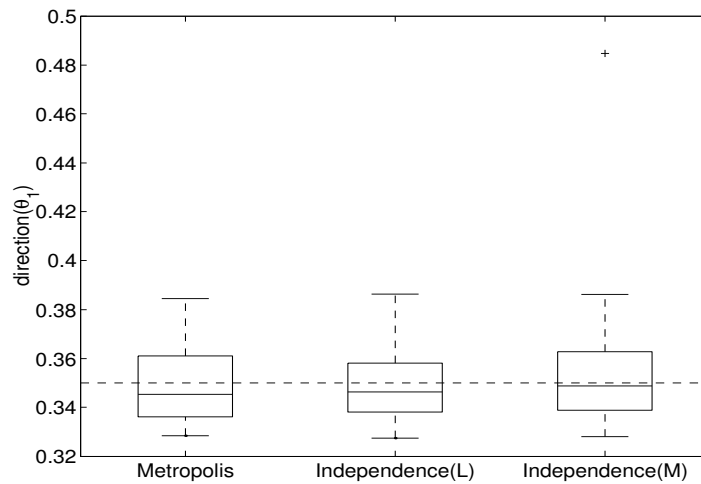


Figure 46. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

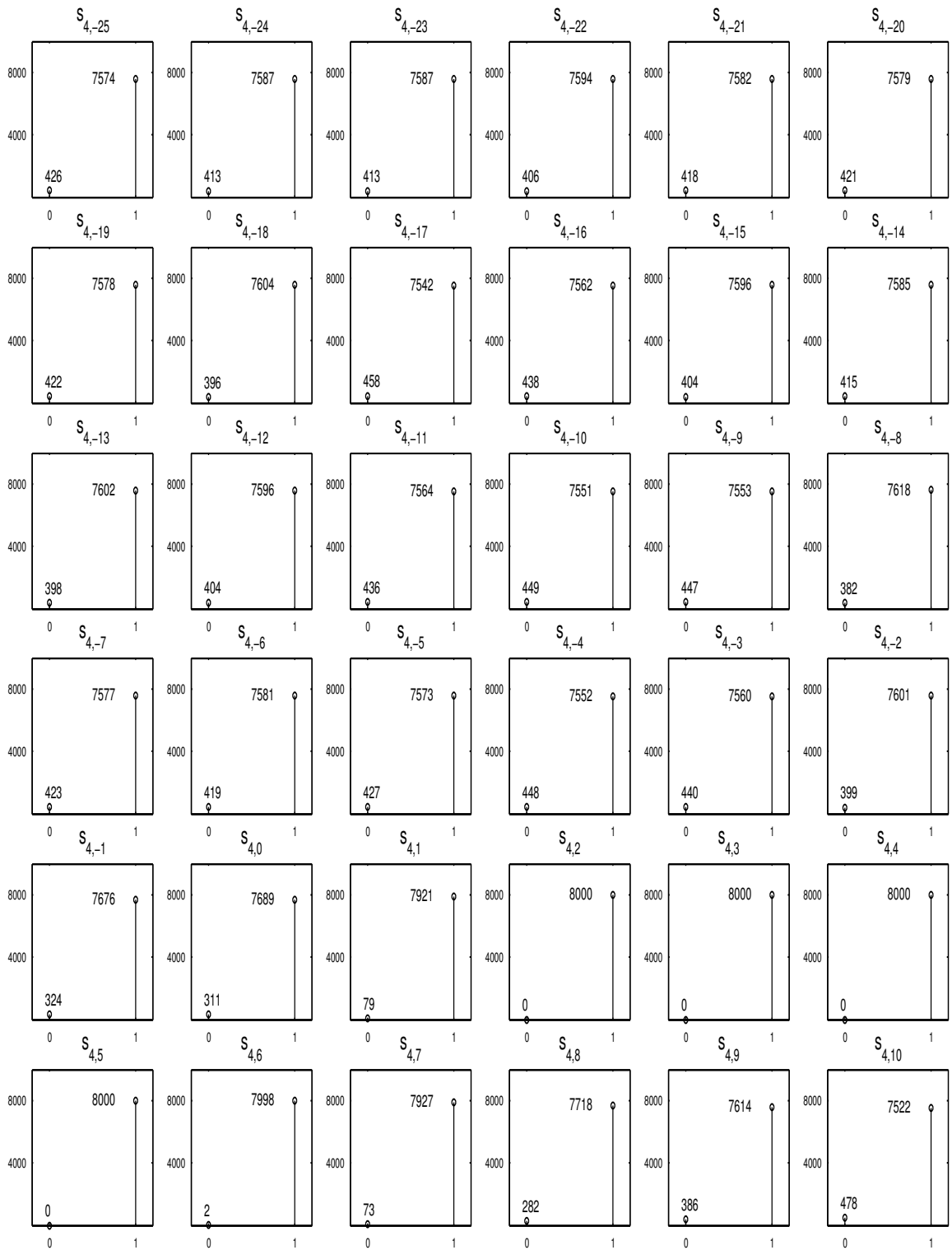


Figure 47. Histograms for the posterior probability of $s_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$, using the metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000.

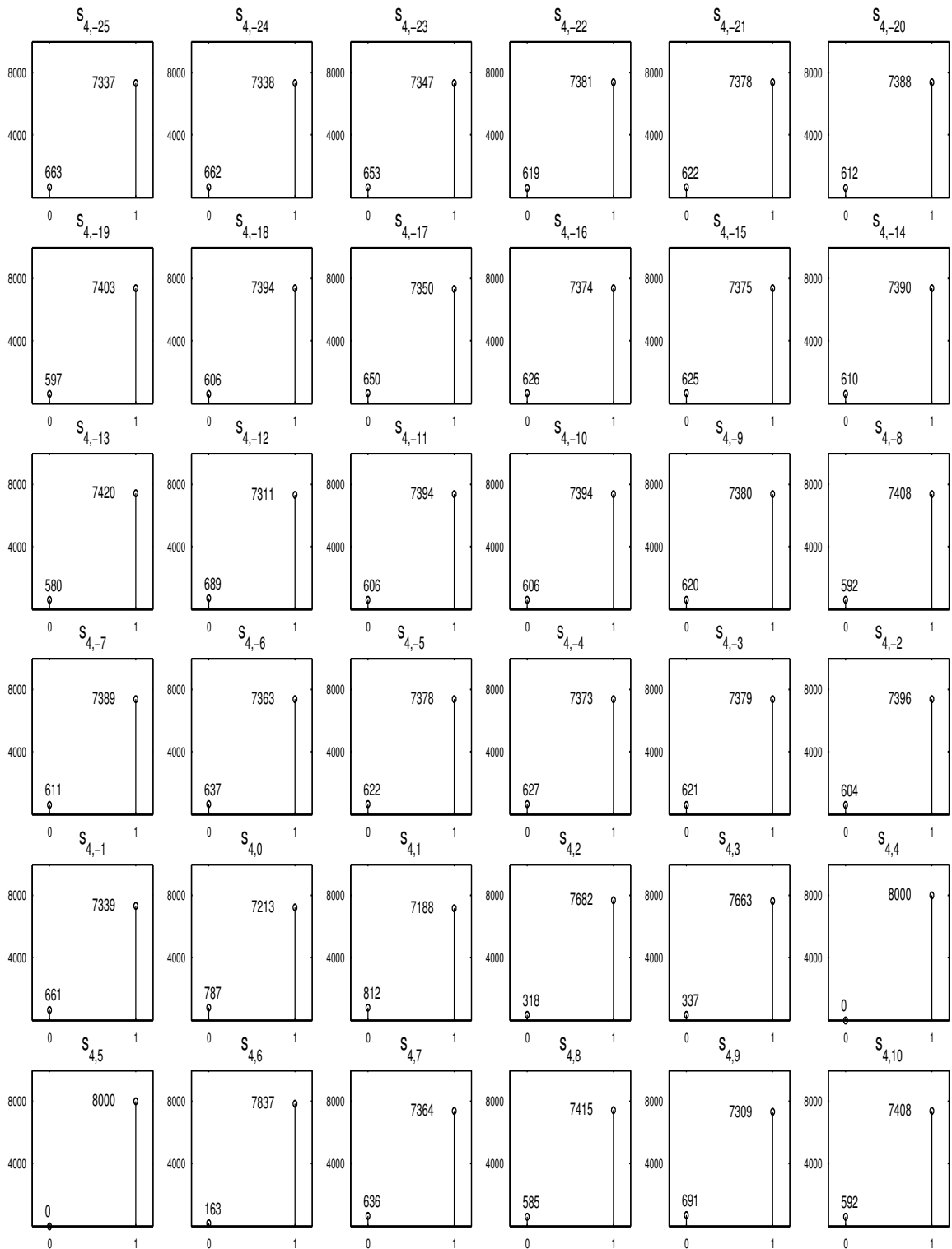


Figure 48. Histograms for the posterior probability of $s_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$, using the metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000.

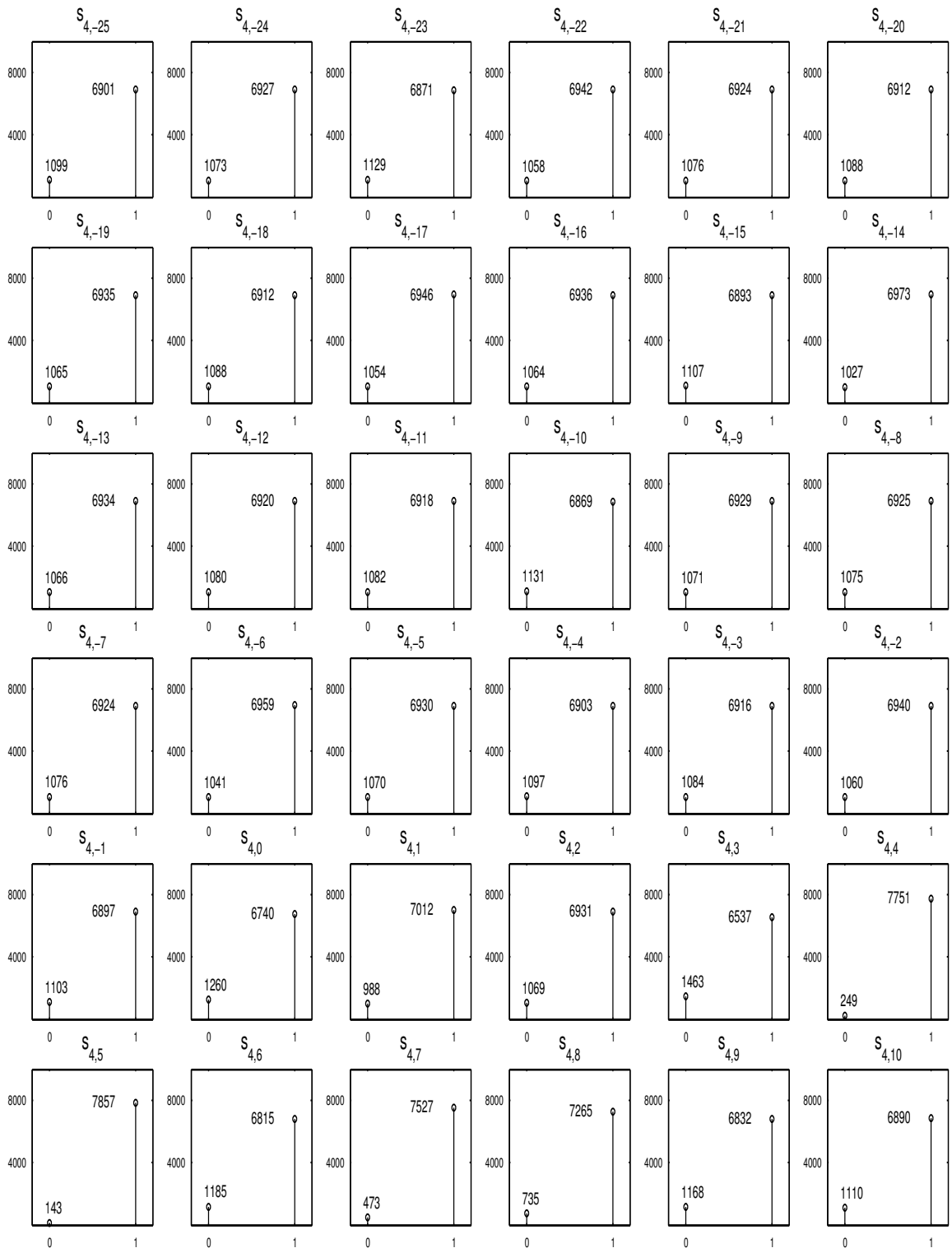


Figure 49. Histograms for the posterior probability of $s_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the maximum support and shrinkage and the maximum resolution $m_0 = 5$, using the metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000.

Table 5. Simulation results for $y_i = 2\sqrt{z_i(1-z_i)} \sin\left(\frac{2.1\pi}{z_i+0.05}\right) + \varepsilon_i$ with the adjusted support and with shrinkage where $z_i = X^i T(\theta)$. The values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined based on an acceptance rate (see Subsection 3.5.2).

Metropolis algorithm, $N(\theta_P, \sigma_1^2)$

resolution	(θ, σ) true value	θ		σ		r	
		bias	mse	bias	mse	bias	mse
$m_0 = 5$	(0.35, 0.02)	-6.36e-3	4.06e-5	1.81e-1	3.26e-2	4.93e-4	1.79e-2
	(0.35, 0.5)	-6.63e-3	7.32e-5	5.45e-2	3.70e-2	9.26e-3	5.27e-2
	(0.35, 1)	3.14e-4	1.71e-4	1.54e-2	2.93e-3	7.95e-3	1.36e-1
$m_0 = 6$	(0.35, 0.02)	-3.41e-4	6.20e-7	3.17e-1	1.01e-2	-1.00e-2	3.36e-4

Independence sampler, $N(\theta_L, \sigma_2^2)$

$m_0 = 5$	(0.35, 0.02)	-6.46e-3	4.19e-5	1.79e-1	3.18e-2	4.63e-4	1.79e-2
	(0.35, 0.5)	-6.25e-3	6.71e-5	5.44e-2	3.68e-3	9.20e-3	5.29e-2
	(0.35, 1)	-2.23e-3	2.55e-4	3.99e-2	4.19e-3	8.92e-3	1.42e-1
$m_0 = 6$	(0.35, 0.02)	2.02e-4	4.92e-7	3.12e-1	9.77e-2	-7.11e-4	3.34e-3

Independence sampler, $N(\theta_M, \sigma_2^2)$

$m_0 = 5$	(0.35, 0.02)	-6.77e-3	4.71e-5	1.77e-1	3.14e-2	4.63e-4	1.79e-2
	(0.35, 0.5)	-6.29e-3	6.80e-5	5.39e-2	3.64e-3	9.05e-3	5.30e-2
	(0.35, 1)	-3.04e-3	2.65e-4	4.15e-2	4.69e-3	8.92e-3	1.42e-1

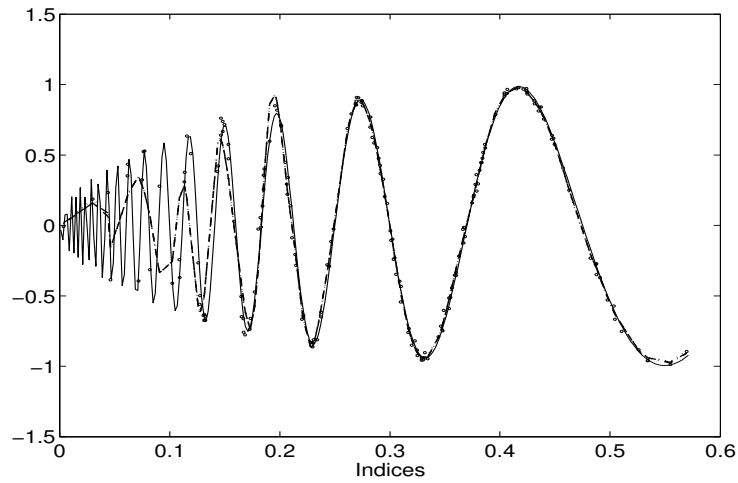


Figure 50. Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence based on two lines(dashed line), and the independence based on the Matlab command `fminsearch`(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$.

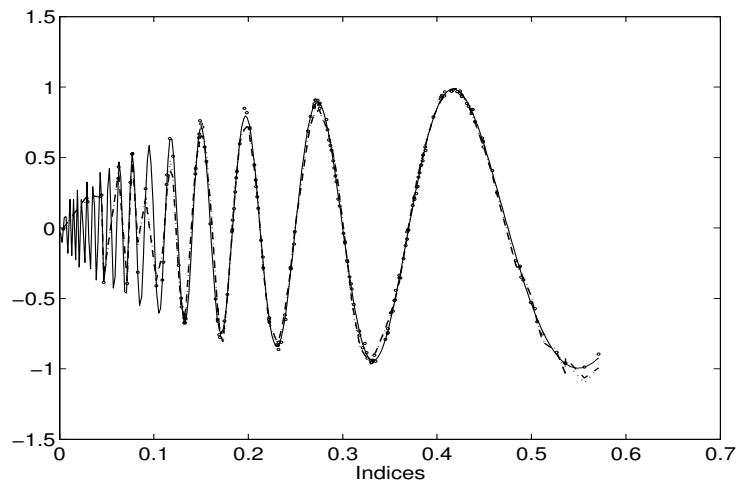


Figure 51. Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence based on two lines(dashed line), and the independence based on the Matlab command `fminsearch`(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 6$.

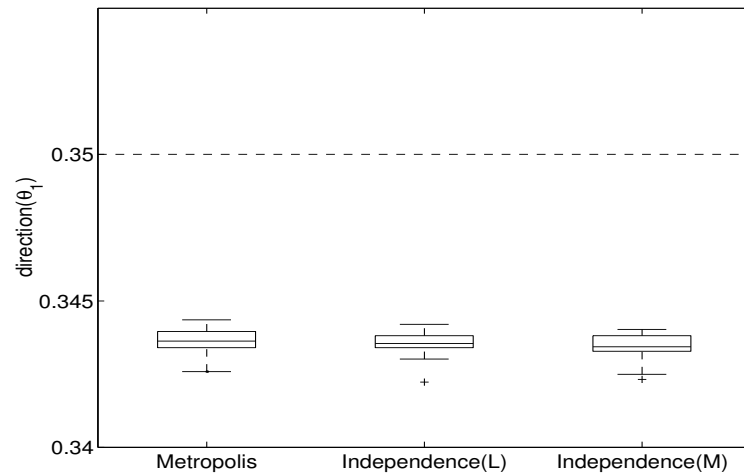


Figure 52. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

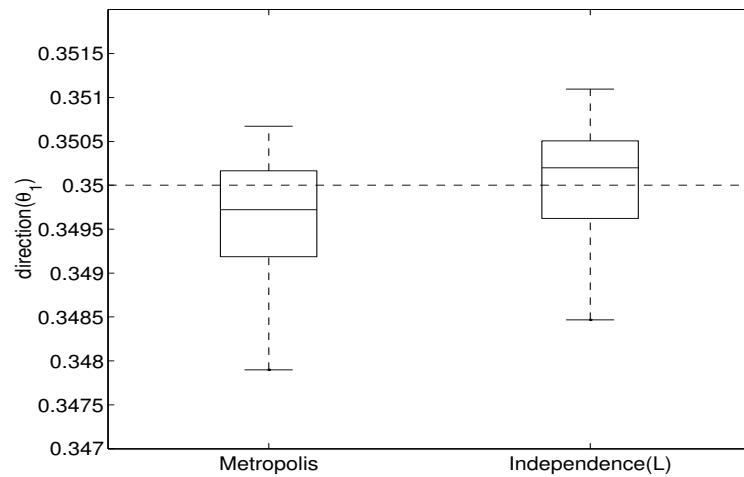


Figure 53. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 6$. The box has lines at the lower quartile, median, and upper quartile values.

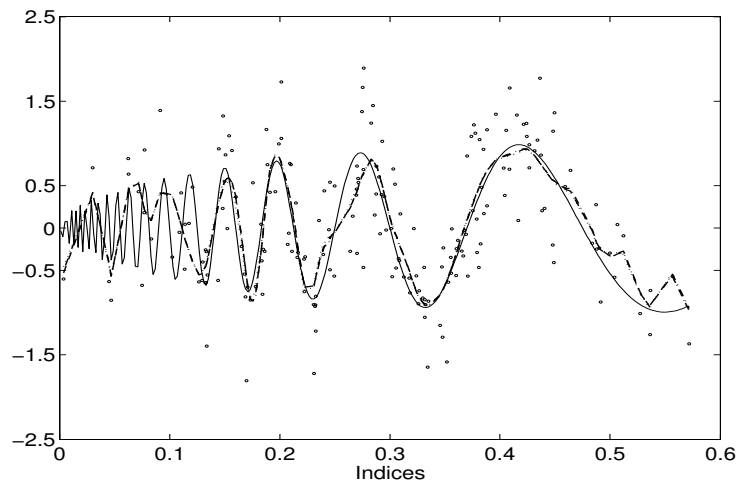


Figure 54. Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence based on two lines(dashed line), and the independence based on the Matlab command `fminsearch`(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$.

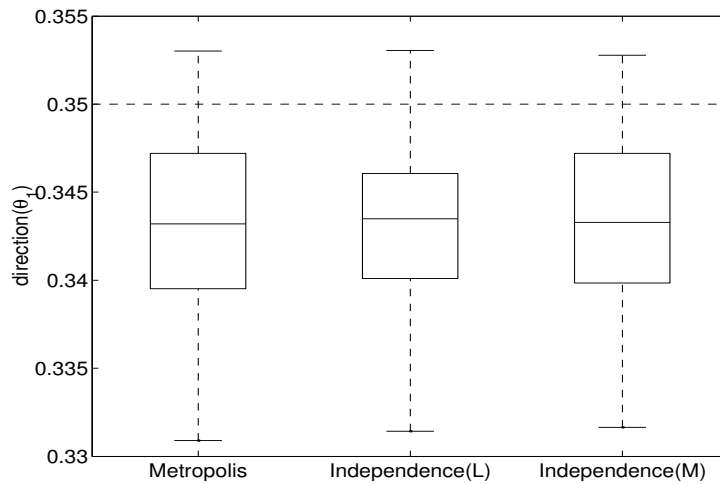


Figure 55. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

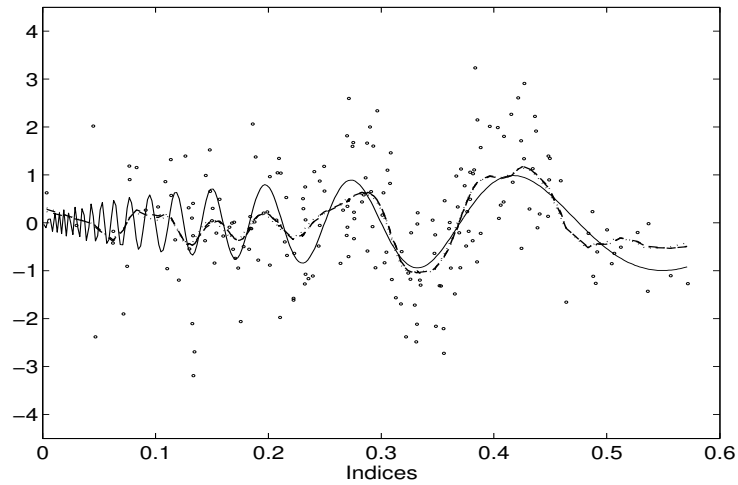


Figure 56. Posterior estimated mean functions for the Metropolis algorithm(dotted line), the independence based on two lines(dashed line), and the independence based on the Matlab command `fminsearch`(dash-dot line) together with the true curve(solid line) on the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$.

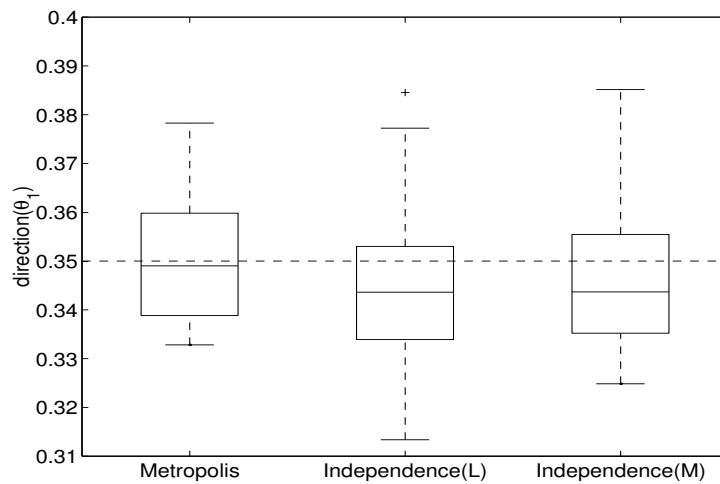


Figure 57. Boxplots for the posterior means of the direction parameter for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$. The box has lines at the lower quartile, median, and upper quartile values.

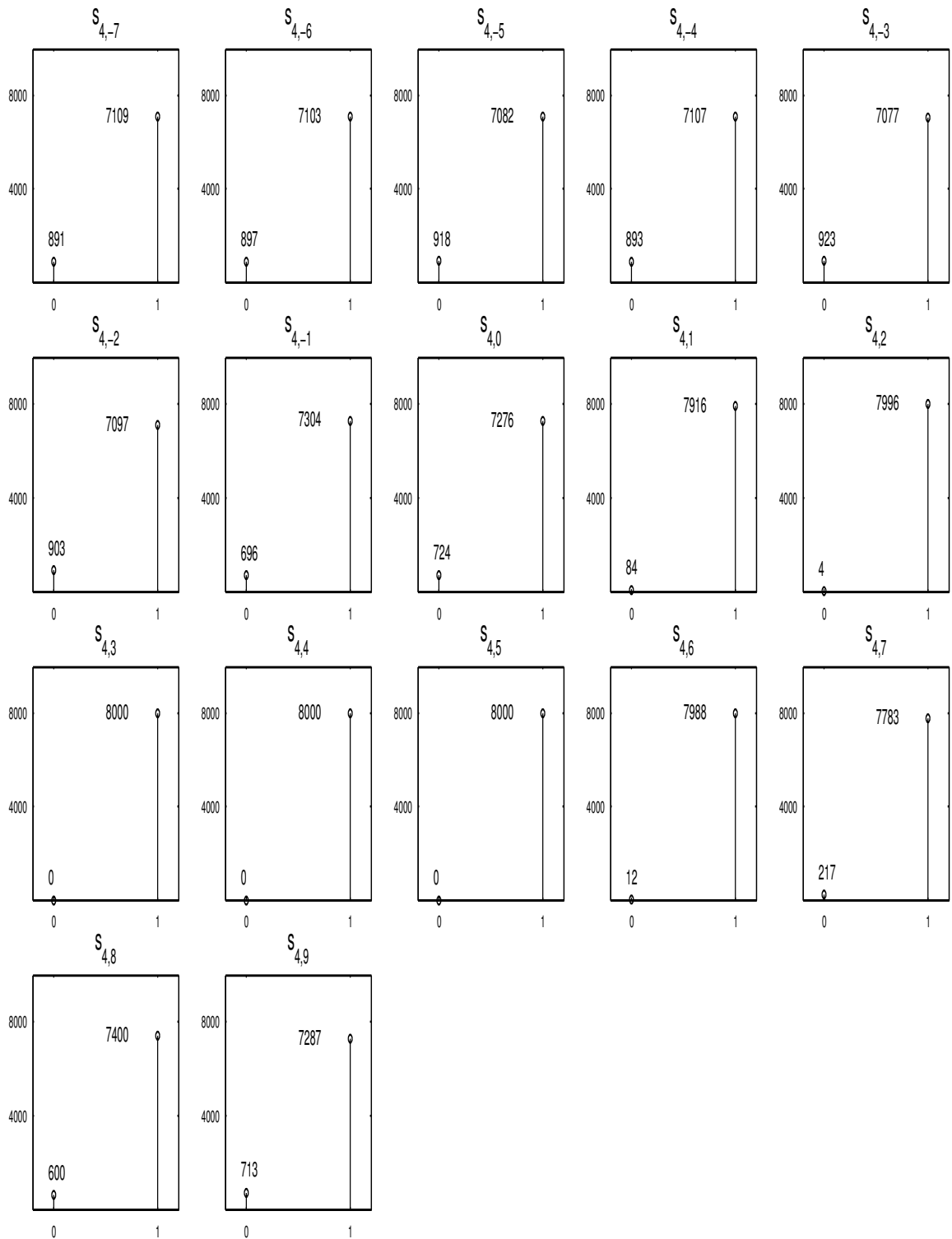


Figure 58. Histograms for the posterior distribution of $w_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.02^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000.

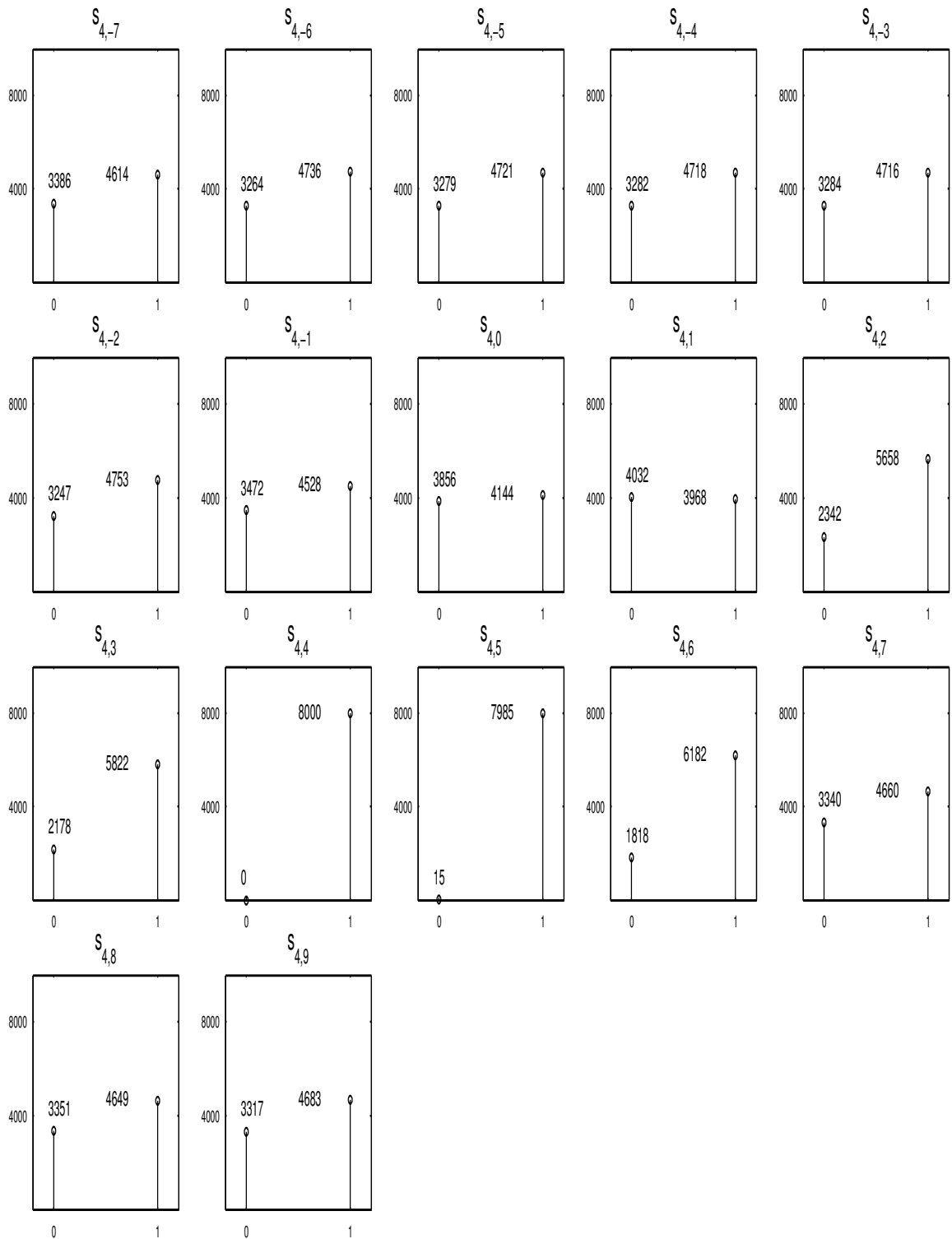


Figure 59. Histograms for the posterior distribution of $w_{4,k}$ for $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 0.5^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000.

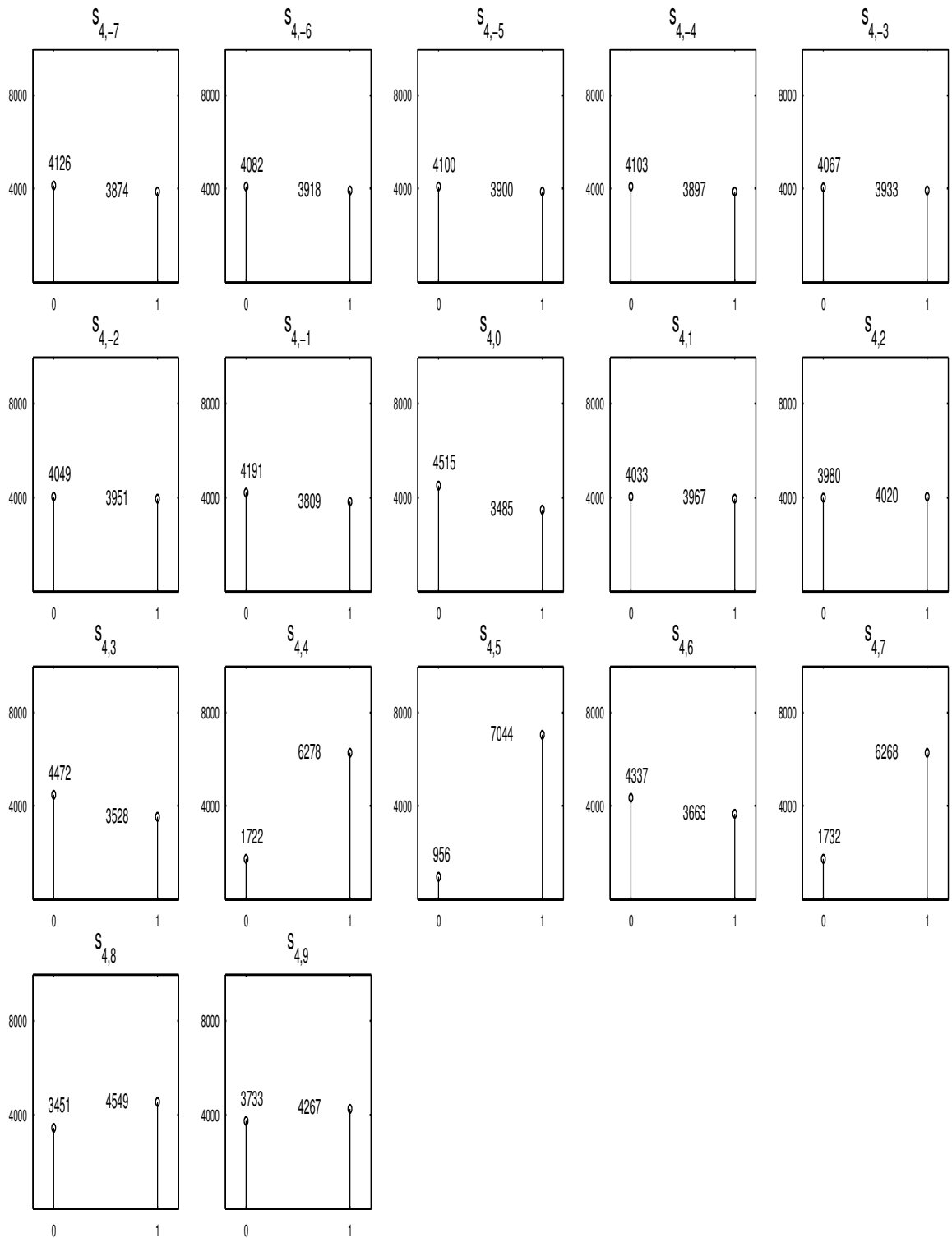


Figure 60. Histograms for the posterior distribution of $w_{4,k}$ for the true direction $\theta = 0.35$ and the Doppler function in Example 2 with $\sigma^2 = 1^2$ for the adjusted support and shrinkage and the maximum resolution $m_0 = 5$, using the Metropolis algorithm. The scale on the y-axis is frequency with the maximum frequency 8000.

4.3 A real data example

We used the data from (Yu and Ruppert, 2002) on the relationship of ozone with the covariates solar radiation, wind speed, and temperature. There are 111 days of observations from May to September 1973 in New York. We apply a single index model using the Bayesian approach and the smooth part from (3.1), since adding the detail parts increased residual sums of squares in the MCMC simulations. Measurements of daily ozone concentration in parts per billion(ppb), solar radiation in Langleys(langleys), wind speed in miles/hour(mph), and daily maximum temperature in degrees Fahrenheit(F).

Table 6 summarizes the results. Figure 61 shows three curve estimates and the box plots of the direction parameter for the Metropolis algorithm and two independence samplers.

Table 6. Results on air pollution data

	Radiation (β_1)		Wind (β_2)		Temperature (β_3)	
	mean	std.	mean	std.	mean	std.
Metropolis	0.0236	0.0072	-0.7895	0.0808	0.5997	0.0994
Independence(L)	0.0236	0.0072	-0.7860	0.0831	0.6036	0.1017
Independence(M)	0.0238	0.0072	-0.7833	0.0877	0.6061	0.1038

We conclude from Table 6 and Figure 61 that there is not a substantial difference between the Metropolis algorithm and the two independence samplers in estimating the parameters and the function in the single index model considered. As shown in Figure 62, the error terms exhibit no obvious pattern and their distribution appears to be approximately normal. The posterior samples of the coefficients $c_{0,k}$ are shown in Figures 63-65.

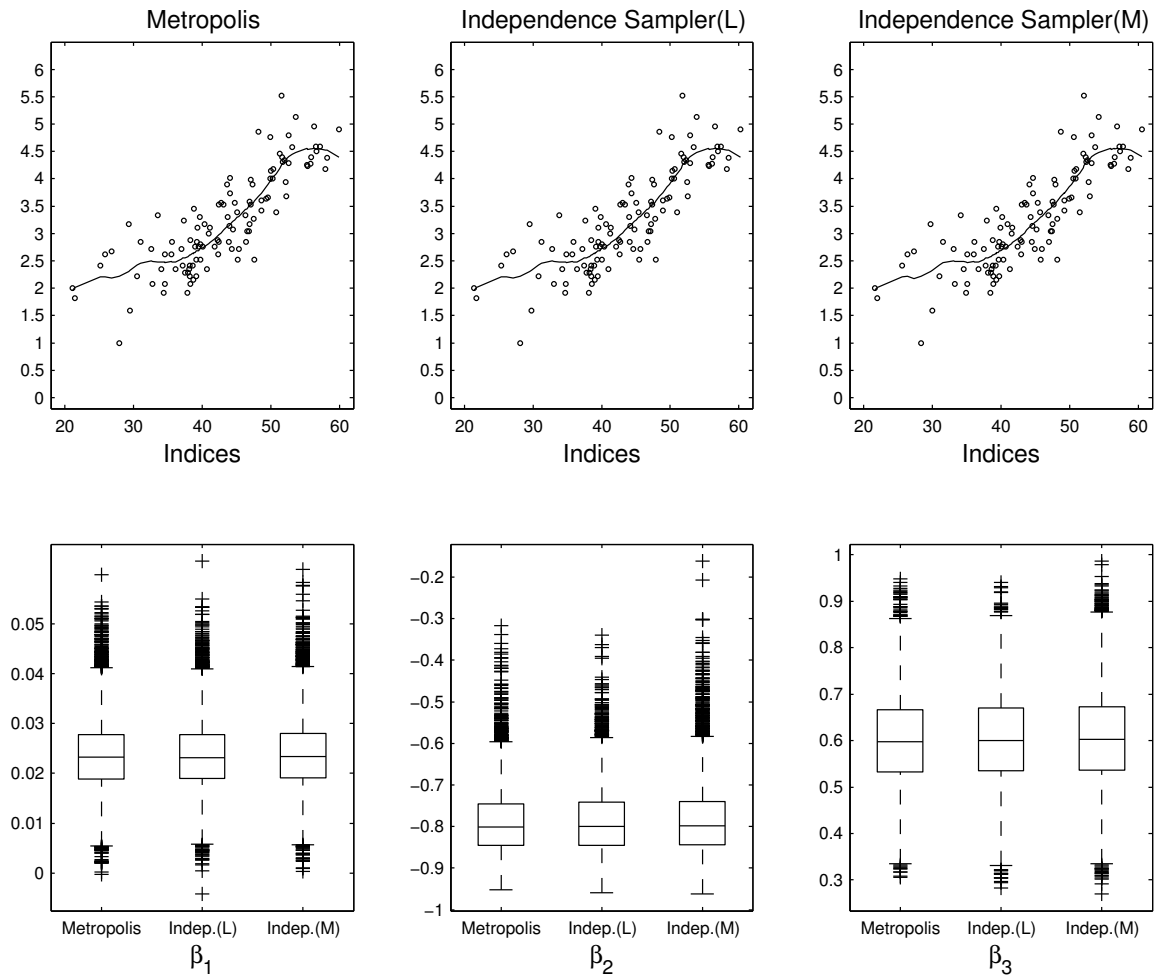


Figure 61. The estimated function and the boxplots of the direction parameter for the Metropolis algorithm and two independence samplers. The box has lines at the lower quartile, median, and upper quartile values.

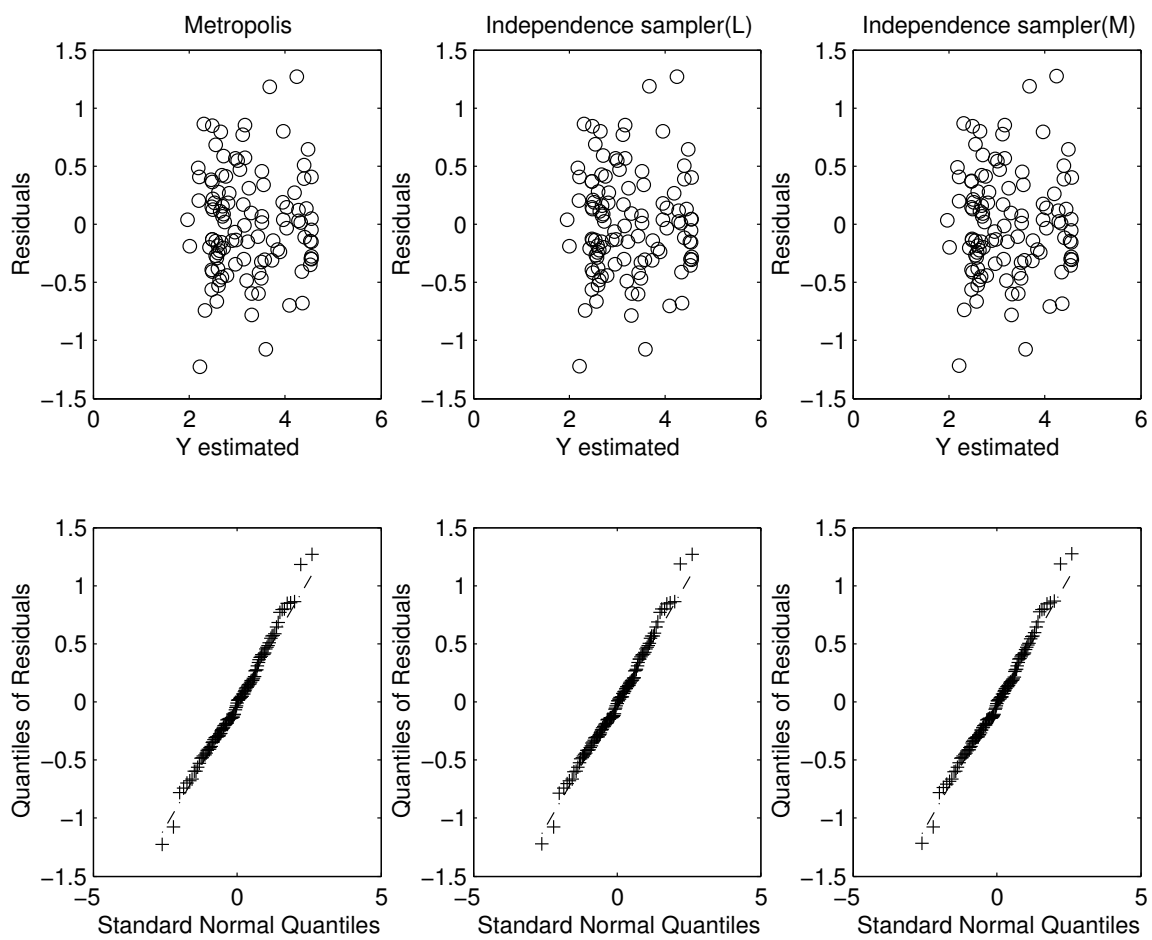


Figure 62. Residuals for the Metropolis algorithm and two independence samplers.

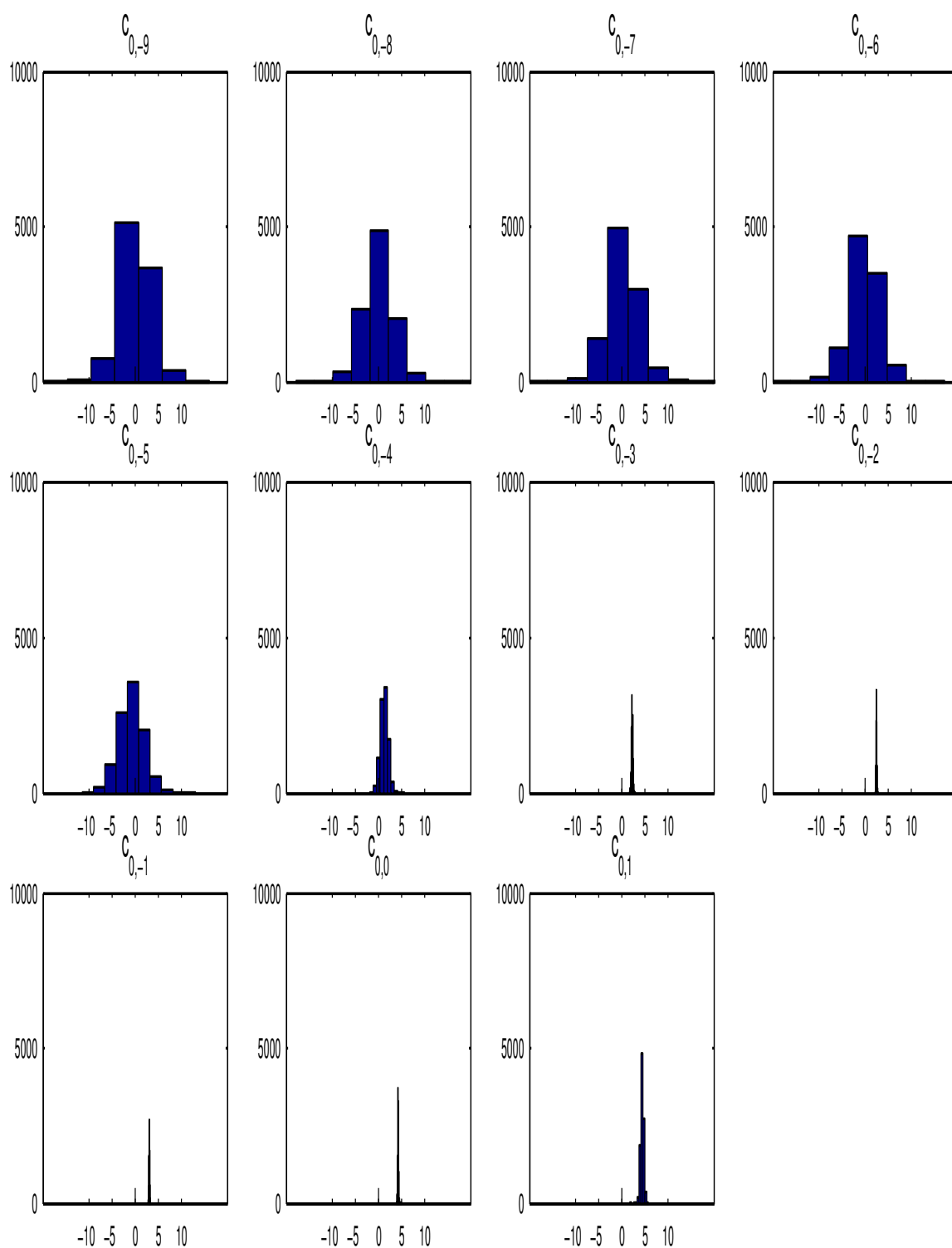


Figure 63. Histograms of the coefficients $c_{0,k}$ for the Metropolis algorithm after the burn-in period. The scale on the y-axis is frequency with the maximum frequency 9000.

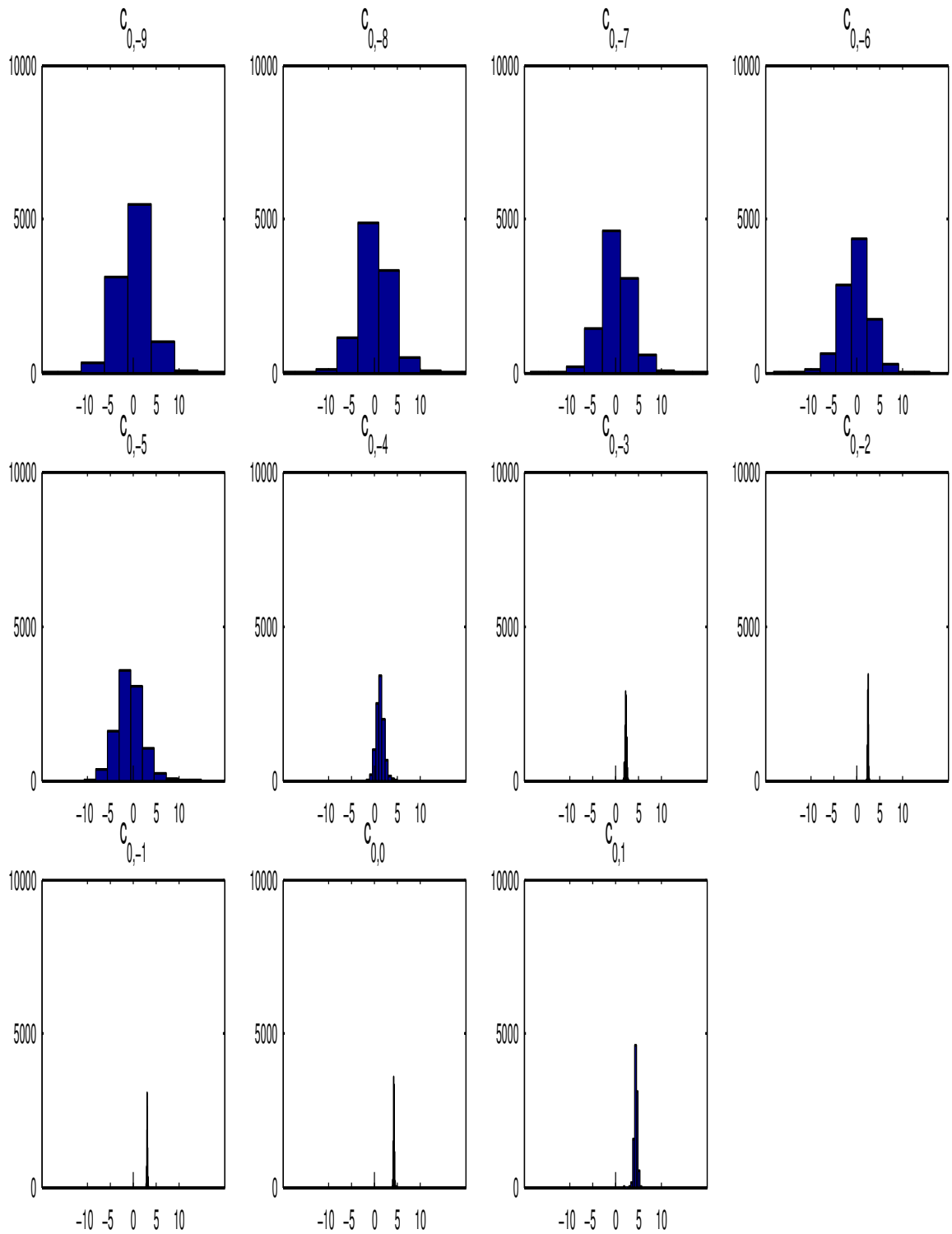


Figure 64. Histograms of the coefficients $c_{0,k}$ for the independence sampler based on two lines after the burn-in period. The scale on the y-axis is frequency with the maximum frequency 9000.

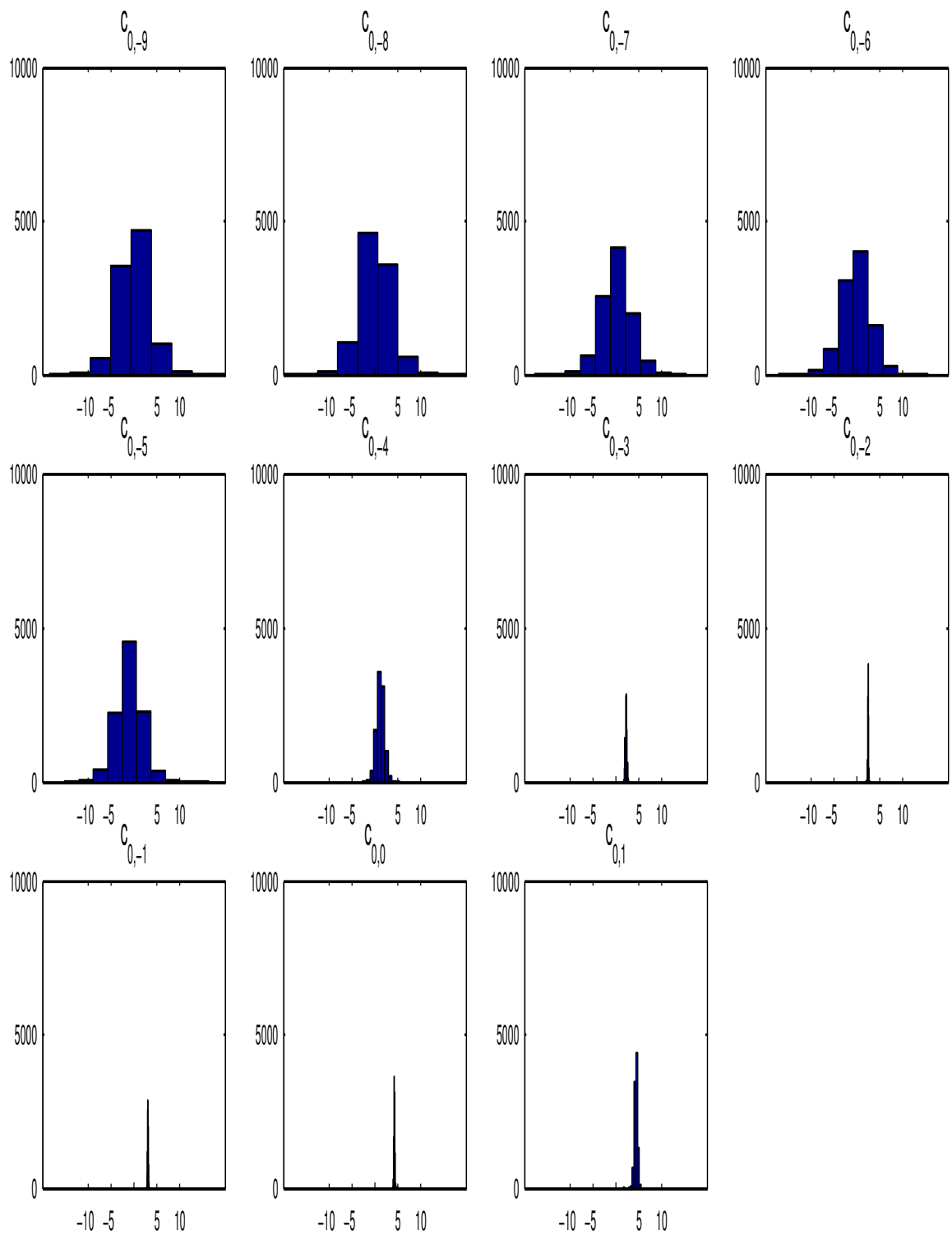


Figure 65. Histograms of the coefficients $c_{0,k}$ for the independence sampler based on the Matlab command `fminsearch` after the burn-in period. The scale on the y-axis is frequency with the maximum frequency 9000.

CHAPTER V

CONCLUSIONS

5.1 Summary

A Bayesian approach to single index models has proposed. Since a function $r(\cdot)$ is unknown, a wavelet series is thought of as a good approximation to any function $r(\cdot)$ in the L^2 space. We have considered two types for representing the function: no shrinkage and shrinkage rules. A hierarchical mixture model for wavelet coefficients is proposed and for the wavelet shrinkage representation a "pseudo prior" is used. Since the direction β is confined to the unit vector, it is convenient to transform β to a spherical polar coordinate θ .

To implement posterior inference we proposed a Metropolis algorithm and an independence sampler for the direction parameter θ , since the posterior distribution is not known. We introduce simple ways to choose variances and a mode (mean) of a proposal distribution for the direction θ . The real data analysis and the Monte Carlo study of a cosine function and a Doppler function with three variances of errors illustrate that the performances of the two samplers for estimating the direction and the two functions are not different, but the Metropolis algorithm is faster than the independence sampler from section (3.4) and the independence sampler based on the matlab command *fminsearch* is slowest.

5.2 Future work

So far the resolution m_0 is fixed. Because of this, from the Monte Carlos study the results of the direction did not include the true value (see Boxplots of the direction for the Doppler function with $\sigma^2 = 0.02^2$). So we require to apply a full posterior inference to single index models, i.e., the resolution is regarded as a parameter.

REFERENCES

- Abramovich, F., Sapatinas, T., and Silverman, B.W. (1998), "Wavelet Thresholding via a Bayesian Approach", *Journal of the Royal Statistical Society Ser. B*, 60(3), 725-749.
- Antoniadis, A., Grégoire, G. and McKeague, I. (1994), "Wavelet Methods for Curve Estimation", *Journal of the American Statistical Association*, 89, 1340-1353.
- Bellman, R. (1961), *Adaptive Control Processes: A Guided Tour*, Princeton, NJ: Princeton University Press.
- Cai, T. and Brown, L. (1997), "Wavelet Shrinkage for Nonequispaced Samples", *The Annals of Statistics*, 26, 1783-1799.
- Carlin, B. P. and Chib, S. (1995), "Bayesian Model Choice via Markov Chain Monte Carlo", *Journal of the Royal Statistical Society Ser. B*, 57, 473-484.
- Carroll, R. J., Fan, J., Gijbels, I., and Wand, M.P. (1997), "Generalized Partially Linear Single-Index Models", *Journal of the American Statistical Association*, 92, 477-489.
- Chipman, H., Kolaczyk, E., and McCulloch, R. (1997), "Adaptive Bayesian Wavelet Shrinkage", *Journal of the American Statistical Association*, 92, 1413-1421.
- Chui, K. (1992), *An Introduction to Wavelets*, Boston, MA: Academic Press.
- Clyde, M., Parmigiani, G., Vidakovic, B. (1998), "Multiple Shrinkage and Subset Selection in Wavelets", *Biometrika*, 85, 391-402.
- Daubechies, Ingrid (1992), *Ten Lectures on Wavelets*, Philadelphia: Society for Industrial and Applied Mathematics (SIAM).
- Hall, P. and Turlach, B.A. (1997). "Interpolation Methods for Nonlinear Wavelet Regression with Irregularly Spaced Design", *The Annals of Statistics*, 25, 1912-1925.
- Härdle, W. , Hall, P., and Ichimura, H. (1993), "Optimal Smoothing in Single-Index Models", *The Annals of Statistics*, 21, 157-178.

- Härdle, W. and Stoker, T. M. (1993), “Investing Smooth Multiple Regression by the Method of Average Derivatives”, *Journal of the American Statistical Association*, 84, 986-995.
- Hart, J. D. (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, New York: Springer-Verlag.
- McCullagh, P. and Nelder, J. A. (1983), *Generalized Linear Models*, London: Chapman and Hall.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953), “Equations of State Calculations by Fast Computing Machines”, *Journal of Chemical Physics*, 21, 1087-1091.
- Meyer, Y. (1992), *Wavelets and Operator*, Cambridge University Press.
- Müller, P. and Vidakovic, B. (1999), “MCMC Methods in Wavelet Shrinkage: Non-Equally Spaced Regression, Density and Spectral Density Estimation”, in *Bayesian Inference in Wavelet Based Models*, eds. P. Müller and B. Vidakovic, New York: Springer-Verlag.
- Sardy, S., Percival, D. B., Bruce, A. G., Gao, H.-Y., and Stuetzle, W. (1997), “Wavelet Denoising for Unequally Spaced Data”, *Statistics and Computing*, 9, 65-75.
- Stoker, T. M. (1986), “Consistent Estimation of Scaled Coefficients”, *Econometrica*, 54, 1461-1481.
- Tierney (1994), “Markov Chains for Exploring Posterior Distributions”, *The Annals of Statistics*, 22, 1701-1762.
- Vidakovic, B. (1998), “Nonlinear Wavelet Shrinkage with Bayes Rules and Bayes Factors”, *Journal of the American Statistical Association*, 93, 173-179.
- Yan, P. and Kohn, R. (1999), “Wavelet Nonparametric Regression Using Basis Averaging”, in *Bayesian Inference in Wavelet Based Models*, eds. P. Müller and B. Vidakovic, New York: Springer-Verlag.
- Yu, Y. and Ruppert, D. (2002), “Penalized Spline Estimation for Partially Linear Single-

Index Models”, *Journal of the American Statistical Association*, 97, 1042-1054.

APPENDIX A

NOTATION AND SOME USEFUL PROPERTIES

A.1 L^2 space

Definition A.1. A function r is in the $L^2[a, b]$ space if r is square integrable, i.e.,

$$\|r\| = \left[\int_a^b |r(z)|^2 dz \right]^{1/2} < \infty. \quad (\text{A-1})$$

A.2 Matrices

Since we use matrices in Appendix A, we now introduce some matrix notation. For notational simplification, matrices are represented by boldface upper case letters, e.g., \mathbf{A} , \mathbf{X} , \mathbf{R} . Their elements are represented by small letters with subscripts. Row(column) vectors are written as capital letters with superscripts(subscripts).

Notation A.1. Let \mathbf{A} denote an $n \times p$ matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} = (a_{ij}), \quad (\text{A-2})$$

where a_{ij} is the element in row i and column j of the matrix \mathbf{A} , $i = 1, \dots, n$; $j = 1, \dots, p$.

Notation A.2. We write the rows(the columns) of the matrix \mathbf{A}

$$\mathbf{A} = [A_1 \ A_2 \ \dots \ A_p] = \begin{bmatrix} A^1 \\ A^2 \\ \vdots \\ A^n \end{bmatrix}, \quad (\text{A-3})$$

where $A^i = [a_{i1} \ \dots \ a_{ip}]$ and $A_j = [a_{1j} \ \dots \ a_{nj}]^T$, $i = 1, \dots, n$; $j = 1, \dots, p$.

Definition A.2. Let \mathbf{A} be a $p \times p$ matrix. Suppose that E is a nonzero vector in \mathbb{R}^p and λ a number (possibly zero) such that

$$\mathbf{A}E = \lambda E. \quad (\text{A-4})$$

Then E is called an eigenvector of \mathbf{A} , and λ is called an eigenvalue of \mathbf{A} .

Theorem A.1. Let \mathbf{A} be a $p \times p$ symmetric matrix. Then the maximum of $\beta^T \mathbf{A} \beta$ subject to

$$\beta^T \beta = \|\beta\|^2 = 1, \quad (\text{A-5})$$

is attained when β is the eigenvector of \mathbf{A} corresponding to the largest eigenvalue of \mathbf{A} . Thus if λ_1 is the largest eigenvalue of \mathbf{A} , then subject to the constraint (A-5),

$$\max_{\beta} \beta^T \mathbf{A} \beta = \lambda_1. \quad (\text{A-6})$$

Proof Let $\mathbf{A} = \mathbf{E} \mathbf{D} \mathbf{E}^T$ be a spectral decomposition of the symmetric matrix \mathbf{A} . Let $Z = \mathbf{E}^T \beta$. Then $Z^T Z = \beta^T \mathbf{E} \mathbf{E}^T \beta = \beta^T \beta$, and what we wish to prove is equivalent to

$$\max_Z Z^T \mathbf{D} Z = \max_Z \sum \lambda_i z_i^2 \quad \text{subject to} \quad Z^T Z = 1. \quad (\text{A-7})$$

If the eigenvalues are written in descending order then (A-7) satisfies

$$\max \sum \lambda_i z_i^2 \leq \lambda_1 \max \sum z_i^2 = \lambda_1.$$

When $\beta = E_1$, then $\beta^T \mathbf{A} \beta = E_1^T \mathbf{E} \mathbf{D} \mathbf{E}^T E_1 = \lambda_1$, where E_1 is the eigenvector of \mathbf{A} corresponding to λ_1 , and hence the maximum of λ_1 is attainable. This complete the proof.

Corollary A.1. The maximum of $|X\beta|$, where X is $(1 \times p)$ and β $(p \times 1)$, subject to (A-5), is

$$[XX^T]^{1/2}. \quad (\text{A-8})$$

Proof Apply Theorem A.1 with $\beta^T \mathbf{A} \beta = (X\beta)^2 = \beta^T (X^T X) \beta$.

APPENDIX B

CALCULATIONS OF POSTERIOR DISTRIBUTIONS FOR NO SHRINKAGE

Let $Z = \mathbf{X}\beta \in \{z_{(1)}, z_{(n)}\}$ be given and the boundaries of the translations k of the coefficients $\{c_{0,k}, w_{j,k}\}$ be denoted as $[a_0, b_0]$ and $[a_j, b_j]$, $j = 0, \dots, m_0$, respectively, where all boundaries are integers. For no shrinkage, let $\Omega_1 = [\{c_{0,k}\}, \{w_{j,k}\}, \tau, \sigma^2, \theta]$ and let $\Omega_1(-\xi)$ denote Ω_1 without the parameter ξ , which could be any of $\{c_{0,k}\}, \{w_{j,k}\}, \tau, \sigma^2$, or θ .

From (3.7), (3.8), and (3.13), the conditional distributions of the parameters Ω_1 are as listed.

- The joint distribution of the data Y and the parameters Ω_1 , conditional on \mathbf{X} and the resolution m_0 , is

$$\begin{aligned}
P(Y, \Omega_1 | \mathbf{X}, m_0) &= \text{Likelihood} \times \text{Joint prior} \\
&= P(Y | \{c_{0,k}\}, \{w_{j,k}\}, \sigma^2, \theta, \mathbf{X}, m_0) \cdot P(\Omega_1) \\
&\propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_3(y_i)\right) \cdot \left[\prod_{k=a_0}^{b_0} \tau^{-\frac{1}{2}} \exp\left(-\frac{c_{0,k}^2}{2\tau}\right) \right] \\
&\quad \times \left[\prod_{j=0}^{m_0} \prod_{k=a_j}^{b_j} \tau^{-\frac{1}{2}} \exp\left(-\frac{w_{j,k}^2}{2^{-(j-1)}\tau}\right) \right] \cdot \tau^{-(a_\tau+1)} \exp\left(-\frac{1}{\tau b_\tau}\right) \\
&\quad \times \sigma^{2-(a_\nu+1)} \exp\left(-\frac{1}{\sigma^2 b_\nu}\right) \\
&\propto (\sigma^2)^{-\left(\frac{n}{2}+a_\nu+1\right)} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_3(y_i) - \frac{1}{\sigma^2 b_\nu}\right) \tau^{-\left(\frac{s}{2}+a_\tau+1\right)} \\
&\quad \times \exp\left(-\frac{1}{\tau b_\tau} - \sum_{k=a_0}^{b_0} \frac{c_{0,k}^2}{2\tau} - \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} \frac{w_{j,k}^2}{2^{-(j-1)}\tau}\right), \tag{B-1}
\end{aligned}$$

where $s = b_0 - a_0 + 1 + \sum_{j=0}^{m_0} (b_j - a_j + 1)$.

- The conditional distribution of σ^2 is

$$\begin{aligned} P(\sigma^2 | \Omega_1(-\sigma^2), Y, \mathbf{X}, m_0) &\propto (\sigma^2)^{-v_1} \exp\left[-\frac{v_2}{\sigma^2}\right] \\ &\propto IG(v_1 - 1, 1/v_2), \end{aligned} \quad (\text{B-2})$$

where $v_1 = \frac{n}{2} + a_v + 1$ and $v_2 = \frac{1}{b_v} + \frac{1}{2} \sum_{i=1}^n Q_3(y_i)$.

- The conditional distribution of τ is

$$\begin{aligned} P(\tau | \Omega_1(-\tau), Y, \mathbf{X}, m_0) &\propto \tau^{-v_3} \exp\left[-\frac{v_4}{\tau}\right] \\ &\propto IG(v_3 - 1, 1/v_4), \end{aligned} \quad (\text{B-3})$$

where $v_3 = \frac{s}{2} + a_\tau + 1$ and $v_4 = \frac{1}{b_\tau} + \frac{1}{2} \sum_{k=a_0}^{b_0} c_{0,k}^2 + \frac{1}{2} \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} \frac{w_{j,k}^2}{2^{-j}}$.

- The conditional distribution of c_{0,k_1} is

$$P(c_{0,k_1} | \Omega_1(-c_{0,k_1}), Y, \mathbf{X}, m_0) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_3(y_i) - \frac{1}{2} \sum_{k=a_0}^{b_0} \frac{c_{0,k}^2}{\tau}\right)$$

where

$$\begin{aligned} &-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_3(y_i) - \frac{1}{2} \sum_{k=a_0}^{b_0} \frac{c_{0,k}^2}{\tau} \\ &= -\frac{1}{2\sigma^2} \sum_{i=1}^n \left(E_{(-k_1)}^i - c_{0,k_1} \phi_{0,k_1}(X^i T(\theta)) \right)^2 - \frac{1}{2\tau} c_{0,k_1}^2 - \text{constant}_1 \\ &= -\frac{1}{2} \left[\frac{1}{\sigma^2} \left(c_{0,k_1}^2 \sum_{i=1}^n \phi_{0,k_1}^2(X^i T(\theta)) - 2c_{0,k_1} \sum_{i=1}^n \phi_{0,k_1}(X^i T(\theta)) E_{(-k_1)}^i \right) \right. \\ &\quad \left. + \frac{1}{\tau} c_{0,k_1}^2 \right] - \text{constant}_2 \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{2} \left[c_{0,k_1}^2 \left(\frac{1}{\sigma^2} \sum_{i=1}^n \phi_{0,k_1}^2(X^i T(\theta)) + \frac{1}{\tau} \right) - \frac{2}{\sigma^2} c_{0,k_1} \sum_{i=1}^n \phi_{0,k_1}(X^i T(\theta)) E_{(-k_1)}^i \right] \\
&\quad - \text{constant}_3 \\
&= -\frac{1}{2\sigma_{k_1}^2} (c_{0,k_1} - \mu_{k_1})^2 - \text{constant}_4,
\end{aligned}$$

and

$$\begin{aligned}
\mu_{k_1} &= \frac{\sigma_{k_1}^2}{\sigma^2} \sum_{i=1}^n \phi_{0,k_1}(X^i T(\theta)) \cdot E_{(-k_1)}^i, \quad \frac{1}{\sigma_{k_1}^2} = \frac{1}{\sigma^2} \sum_{i=1}^n \phi_{0,k_1}^2(X^i T(\theta)) + \frac{1}{\tau}, \quad \text{and } E_{(-k_1)}^i = y_i - \\
&\quad \sum_{\substack{k=a_0 \\ k \neq k_1}}^{b_0} c_{0,k} \phi_{0,k}(X^i T(\theta)) - \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} w_{j,k} \psi_{j,k}(X^i T(\theta)) \text{ for } k_1 \in [a_0, b_0].
\end{aligned}$$

Therefore

$$P(c_{0,k_1} | \Omega_1(-c_{0,k_1}), Y, \mathbf{X}, m_0) = N(\mu_{k_1}, \sigma_{k_1}^2). \quad (\text{B-4})$$

- The conditional distribution of w_{j_2, k_2} is

$$P(w_{j_2, k_2} | \Omega_1(-w_{j_2, k_2}), Y, \mathbf{X}, m_0) \propto \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_3(y_i) - \frac{1}{2} \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} \frac{w_{j,k}^2}{2^{-j\tau}} \right)$$

where

$$\begin{aligned}
&-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_3(y_i) - \frac{1}{2} \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} \frac{w_{j,k}^2}{2^{-j\tau}} \\
&= -\frac{1}{2\sigma^2} \sum_{i=1}^n \left(E_{(-j_2, k_2)}^i - w_{j_2, k_2} \psi_{j_2, k_2}(X^i T(\theta)) \right)^2 - \frac{1}{2^{-(j_2-1)\tau}} w_{j_2, k_2}^2 - \text{constant}_1 \\
&= -\frac{1}{2} \left[\frac{1}{\sigma^2} \left(w_{j_2, k_2}^2 \sum_{i=1}^n \psi_{j_2, k_2}^2(X^i T(\theta)) - 2w_{j_2, k_2} \sum_{i=1}^n \psi_{j_2, k_2}(X^i T(\theta)) E_{(-j_2, k_2)}^i \right) \right. \\
&\quad \left. + \frac{1}{2^{-j_2\tau}} w_{j_2, k_2}^2 \right] - \text{constant}_2 \\
&= -\frac{1}{2} \left[w_{j_2, k_2}^2 \left(\frac{1}{\sigma^2} \sum_{i=1}^n \psi_{j_2, k_2}^2(X^i T(\theta)) + \frac{1}{2^{-j_2\tau}} \right) \right.
\end{aligned}$$

$$\begin{aligned}
& \left. -\frac{2}{\sigma^2} w_{j_2, k_2} \sum_{i=1}^n \Psi_{j_2, k_2}(X^i T(\theta)) E_{(-j_2, k_2)}^i \right] - \text{constant}_3 \\
& = -\frac{1}{2\sigma_{(j_2, k_2)}^2} (w_{j_2, k_1} - \mu_{(j_2, k_2)})^2 - \text{constant}_4,
\end{aligned}$$

and

$$\begin{aligned}
\mu_{(j_2, k_2)} &= \frac{\sigma_{(j_2, k_2)}^2}{\sigma^2} \sum_{i=1}^n \Psi_{j_2, k_2}(X^i T(\theta)) \cdot E_{(-j_2, k_1)}^i, \quad \frac{1}{\sigma_{(j_2, k_2)}^2} = \frac{1}{\sigma^2} \sum_{i=1}^n \Psi_{j_2, k_2}^2(X^i T(\theta)) + \frac{1}{2^{-j_2} \tau}, \text{ and} \\
E_{(-j_2, k_2)}^i &= y_i - \sum_{k=a_0}^{b_0} c_{0,k} \Phi_{0,k}(X^i T(\theta)) - \sum_{\substack{j=0 \\ j \neq j_2}}^{m_0} \sum_{\substack{k=a_j \\ k \neq k_2}}^{b_j} w_{j,k} \Psi_{j,k}(X^i T(\theta)) \text{ for } j_2 \in [0, m_0] \text{ and} \\
& k_2 \in [a_j, b_j].
\end{aligned}$$

Therefore

$$P(w_{j_2, k_2} | \Omega_1(-w_{j_2, k_2}), Y, \mathbf{X}, m_0) = N(\mu_{j_2, k_2}, \sigma_{j_2, k_2}^2). \quad (\text{B-5})$$

- The conditional distribution of θ on the $p-1$ dimensional unit sphere is

$$P(\theta | \Omega_1(-\theta), Y, \mathbf{X}, m_0) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_3(y_i)\right). \quad (\text{B-6})$$

Since $r(Z)$ is in L^2 , given $Z \in [z_{(1)}, z_{(n)}]$, there are constants C_1, C_2 such that

$$\begin{aligned}
[r(X^i T(\theta^*)) - \hat{r}(X^i T(\theta))]^2 &\geq C_1 [X^i T(\theta^*) - X^i T(\theta)]^2 \\
&\geq C_2 \|\theta - \theta^*\|^2,
\end{aligned}$$

where θ^* is a mode (mean) direction and \hat{r} is estimated from the MCMC scheme.

Hence there are c_1, \dots, c_{p-1} such that

$$\begin{aligned}
P(\theta | \Omega_1(-\theta), Y, \mathbf{X}, m_0) &\leq \exp\left[-\frac{1}{2\sigma^2} (c_1 (\theta_1 - \theta_1^*)^2 + \dots \right. \\
&\quad \left. + c_{p-1} (\theta_{p-1} - \theta_{p-1}^*)^2)\right]. \quad (\text{B-7})
\end{aligned}$$

APPENDIX C

CALCULATIONS OF POSTERIOR DISTRIBUTIONS FOR WAVELET SHRINKAGE

For wavelet shrinkage, let $\Omega_2 = [\{c_{0,k}\}, \{w_{j,k}\}, \alpha, \tau, \sigma^2, \theta]$. From (3.7), (3.9), and (3.19), the conditional distributions of the parameters Ω_2 are as listed.

- The joint distribution of the data Y and the parameters Ω_2 , conditional on \mathbf{X} and the resolution m_0 , is

$$\begin{aligned}
P(Y, \Omega_2 | \mathbf{X}, m_0) &= \text{Likelihood} \times \text{Joint prior} \\
&= P(Y | \{c_{0,k}\}, \{s_{j,k}\}, \{w_{j,k}\}, \sigma^2, \theta, \mathbf{X}, m_0) \cdot P(\Omega_2) \\
&\propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(y_i)\right) \cdot (\sigma^2)^{-(a_v+1)} \exp\left(-\frac{1}{\sigma^2 b_v}\right) \\
&\quad \times \tau^{-(a_\tau+1)} \exp\left(-\frac{1}{\tau b_\tau}\right) \cdot \left[\prod_{k=a_0}^{b_0} \tau^{-\frac{1}{2}} \exp\left(-\frac{c_{0,k}^2}{2\tau}\right) \right] \\
&\quad \times \left[\prod_{\substack{j=0 \\ j,k \in \{s_{j,k}=1\}}}^{m_0} \prod_{k=a_j}^{b_j} \tau^{-\frac{1}{2}} \exp\left(-\frac{w_{j,k}^2}{2^{-(j-1)}\tau}\right) \right] \cdot \left[\prod_{\substack{j=1 \\ j,k \in \{s_{j,k}=0\}}}^{m_0} \prod_{k=a_j}^{b_j} h(w_{j,k}) \right] \\
&\quad \times \left[\prod_{j=1}^{m_0} \prod_{k=a_j}^{b_j} (\alpha^j)^{s_{j,k}} (1-\alpha^j)^{1-s_{j,k}} \right] \cdot \alpha^{(a_\alpha-1)} (1-\alpha)^{(b_\alpha-1)} \\
&\propto (\sigma^2)^{-\left(\frac{n}{2}+a_v+1\right)} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(y_i) - \frac{1}{\sigma^2 b_v}\right) \tau^{-\left(\frac{s}{2}+a_\tau+1\right)}
\end{aligned}$$

$$\begin{aligned}
& \times \exp \left(-\frac{1}{\tau b_\tau} - \sum_{k=a_0}^{b_0} \frac{c_{0,k}^2}{2\tau} - \underbrace{\sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} \frac{w_{j,k}^2}{2^{-(j-1)\tau}}}_{j,k \in \{s_{j,k}=1\}} \right) \\
& \times \left[\prod_{j=1}^{m_0} \prod_{k=a_j}^{b_j} (\alpha^j)^{s_{j,k}} (1-\alpha^j)^{1-s_{j,k}} \right] \cdot \alpha^{(a_\alpha-1)} (1-\alpha)^{(b_\alpha-1)} \\
& \times \left[\underbrace{\prod_{j=1}^{m_0} \prod_{k=a_j}^{b_j} h(w_{j,k})}_{j,k \in \{s_{j,k}=0\}} \right], \tag{C-1}
\end{aligned}$$

where $Q_4(y_i) = \left[y_i - \sum_{k=a_0}^{b_0} c_{0,k} \phi_{0,k}(X^i G(\theta)) - \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} s_{j,k} w_{j,k} \psi_{j,k}(X^i G(\theta)) \right]^2$ and $s = b_0 - a_0 + 1 + \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} s_{j,k}$.

- The conditional distribution of σ^2 is

$$\begin{aligned}
P(\sigma^2 | \Omega_1(-\sigma^2), Y, \mathbf{X}, m_0) & \propto (\sigma^2)^{-v_1} \exp \left[-\frac{v_2}{\sigma^2} \right] \\
& \propto IG(v_1 - 1, 1/v_2), \tag{C-2}
\end{aligned}$$

where $v_1 = \frac{n}{2} + a_v + 1$ and $v_2 = \frac{1}{b_v} + \frac{1}{2} \sum_{i=1}^n Q_4(y_i)$.

- The conditional distribution of τ is

$$\begin{aligned}
P(\tau | \Omega_1(-\tau), Y, \mathbf{X}, m_0) & \propto \tau^{-v_3} \exp \left[-\frac{v_4}{\tau} \right] \\
& \propto IG(v_3 - 1, 1/v_4), \tag{C-3}
\end{aligned}$$

where $v_3 = \frac{s}{2} + a_\tau + 1$ and $v_4 = \frac{1}{b_\tau} + \frac{1}{2} \sum_{k=a_0}^{b_0} c_{0,k}^2 + \frac{1}{2} \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} \frac{s_{j,k} w_{j,k}^2}{2^{-j}}$.

- The conditional distribution of c_{0,k_1} is

$$P(c_{0,k_1} | \Omega_1(-c_{0,k_1}), Y, \mathbf{X}, m_0) \propto \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(y_i) - \frac{1}{2} \sum_{k=a_0}^{b_0} \frac{c_{0,k}^2}{\tau} \right)$$

where

$$\begin{aligned}
& -\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(y_i) - \frac{1}{2} \sum_{a_0}^{b_0} \frac{c_{0,k}^2}{\tau} \\
&= -\frac{1}{2\sigma^2} \sum_{i=1}^n \left(E_{(-k_1)}^i - c_{0,k_1} \phi_{0,k_1}(X^i T(\theta)) \right)^2 - \frac{1}{2\tau} c_{0,k_1}^2 - \text{constant}_1 \\
&= -\frac{1}{2} \left[\frac{1}{\sigma^2} \left(c_{0,k_1}^2 \sum_{i=1}^n \phi_{0,k_1}^2(X^i T(\theta)) - 2c_{0,k_1} \sum_{i=1}^n \phi_{0,k_1}(X^i T(\theta)) E_{(-k_1)}^i \right) \right. \\
&\quad \left. + \frac{1}{\tau} c_{0,k_1}^2 \right] - \text{constant}_2 \\
&= -\frac{1}{2} \left[c_{0,k_1}^2 \left(\frac{1}{\sigma^2} \sum_{i=1}^n \phi_{0,k_1}^2(X^i T(\theta)) + \frac{1}{\tau} \right) - \frac{2}{\sigma^2} c_{0,k_1} \sum_{i=1}^n \phi_{0,k_1}(X^i T(\theta)) E_{(-k_1)}^i \right] \\
&\quad - \text{constant}_3 \\
&= -\frac{1}{2\sigma_{k_1}^2} (c_{0,k_1} - \mu_{k_1})^2 - \text{constant}_4,
\end{aligned}$$

and

$$\begin{aligned}
\mu_{k_1} &= \frac{\sigma_{k_1}^2}{\sigma^2} \sum_{i=1}^n \phi_{0,k_1}(X^i T(\theta)) \cdot E_{(-k_1)}^i, \quad \frac{1}{\sigma_{k_1}^2} = \frac{1}{\sigma^2} \sum_{i=1}^n \phi_{0,k_1}^2(X^i T(\theta)) + \frac{1}{\tau}, \quad \text{and } E_{(-k_1)}^i = y_i - \\
&\quad \sum_{\substack{k=a_0 \\ k \neq k_1}}^{b_0} c_{0,k} \phi_{0,k}(X^i T(\theta)) - \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} s_{j,k} w_{j,k} \Psi_{j,k}(X^i T(\theta)) \text{ for } k_1 \in [a_0, b_0].
\end{aligned}$$

Therefore

$$P(c_{0,k_1} | \Omega_1(-c_{0,k_1}), Y, \mathbf{X}, m_0) \propto N(\mu_{k_1}, \sigma_{k_1}^2). \quad (\text{C-4})$$

- The conditional distribution of $w_{j_2, k_2} | s_{j_2, k_2} = 1$ is

$$\begin{aligned}
P(w_{j_2, k_2} | \Omega_1(-w_{j_2, k_2}), s_{j_2, k_2} = 1, Y, \mathbf{X}, m_0) &\propto \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(y_i) \right. \\
&\quad \left. - \frac{1}{2} \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} \frac{w_{j,k}^2}{2^{-j}\tau} \right)
\end{aligned}$$

where

$$\begin{aligned}
& -\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(y_i) - \frac{1}{2} \sum_{j=0}^{m_0} \sum_{k=a_j}^{b_j} \frac{w_{j,k}^2}{2^{-j\tau}} \\
&= -\frac{1}{2\sigma^2} \sum_{i=1}^n \left(E_{(-j_2, k_2)}^i - w_{j_2, k_2} \Psi_{j_2, k_2}(X^i T(\theta)) \right)^2 - \frac{1}{2^{-(j_2-1)\tau}} w_{j_2, k_2}^2 - \text{constant}_1 \\
&= -\frac{1}{2} \left[\frac{1}{\sigma^2} \left(w_{j_2, k_2}^2 \sum_{i=1}^n \Psi_{j_2, k_2}^2(X^i T(\theta)) - 2w_{j_2, k_2} \sum_{i=1}^n \Psi_{j_2, k_2}(X^i T(\theta)) E_{(-j_2, k_2)}^i \right) \right. \\
&\quad \left. + \frac{1}{2^{-j_2\tau}} w_{j_2, k_2}^2 \right] - \text{constant}_2 \\
&= -\frac{1}{2} \left[w_{j_2, k_2}^2 \left(\frac{1}{\sigma^2} \sum_{i=1}^n \Psi_{j_2, k_2}^2(X^i T(\theta)) + \frac{1}{2^{-j_2\tau}} \right) \right. \\
&\quad \left. - \frac{2}{\sigma^2} w_{j_2, k_2} \sum_{i=1}^n \Psi_{j_2, k_2}(X^i T(\theta)) E_{(-j_2, k_2)}^i \right] - \text{constant}_3 \\
&= -\frac{1}{2\sigma_{(j_2, k_2)}^2} (w_{j_2, k_2} - \mu_{(j_2, k_2)})^2 - \text{constant}_4,
\end{aligned}$$

and

$$\begin{aligned}
\mu_{(j_2, k_2)} &= \frac{\sigma_{(j_2, k_2)}^2}{\sigma^2} \sum_{i=1}^n \Psi_{j_2, k_2}(X^i T(\theta)) \cdot E_{(-j_2, k_2)}^i, \quad \frac{1}{\sigma_{(j_2, k_2)}^2} = \frac{1}{\sigma^2} \sum_{i=1}^n \Psi_{j_2, k_2}^2(X^i T(\theta)) + \frac{1}{2^{-j_2\tau}}, \text{ and} \\
E_{(-j_2, k_2)}^i &= y_i - \sum_{k=a_0}^{b_0} c_{0,k} \phi_{0,k}(X^i T(\theta)) - \sum_{\substack{j=0 \\ j \neq j_2}}^{m_0} \sum_{\substack{k=a_j \\ k \neq k_2}}^{b_j} s_{j,k} w_{j,k} \Psi_{j,k}(X^i T(\theta)) \text{ for } j_2 \in [0, m_0] \text{ and} \\
& k_2 \in [a_{j_2}, b_{j_2}].
\end{aligned}$$

Therefore

$$P(w_{j_2, k_2} | \Omega_1(-w_{j_2, k_2}), Y, \mathbf{X}, m_0) \propto N(\mu_{j_2, k_2}, \sigma_{j_2, k_2}^2). \quad (\text{C-5})$$

- The conditional distribution of $w_{j_2, k_2} | s_{j_2, k_2} = 0$ is

$$P(w_{j_2, k_2} | \Omega_1(-w_{j_2, k_2}), s_{j_2, k_2} = 0, Y, \mathbf{X}, m_0) \propto h(w_{j_2, k_2}). \quad (\text{C-6})$$

where $h(w_{j_2, k_2})$ is defined by (3.30).

- The conditional probability of α is

$$P(\alpha | \Omega_2(-\alpha), Y, \mathbf{X}, m_0) \propto \left[\prod_{j=1}^{m_0} \prod_{k=a_j}^{b_j} (\alpha^j)^{s_{j,k}} (1-\alpha^j)^{1-s_{j,k}} \right] \times \alpha^{(a\alpha-1)} (1-\alpha)^{(b\alpha-1)} \quad (\text{C-7})$$

- The conditional probability of s_{j_2, k_2} is

$$P(s_{j_2, k_2} | \Omega_2(-s_{j_2, k_2}), Y, \mathbf{X}, m_0) \propto \begin{cases} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(Y_i)\right) (1-\alpha^{j_2}) h(w_{j_2, k_2}), & s_{j_2, k_2} = 0, \\ \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(Y_i)\right) (\alpha^{j_2}) \exp\left(-\frac{w_{j_2, k_2}^2}{22^{-j_2} \tau}\right), & s_{j_2, k_2} = 1, \end{cases} \quad (\text{C-8})$$

for $j_2 \in [1, m_0]$ and $k_2 \in [a_j, b_j]$.

- The conditional distribution of θ on the $p-1$ dimensional unit sphere is

$$P(\theta | \Omega_1(-\theta), Y, \mathbf{X}, m_0) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n Q_4(y_i)\right). \quad (\text{C-9})$$

VITA

Chun Gun Park, son of Chun Woo Park and Jung Ja Yun, was born on October 10, 1968, in Seoul, Korea. He received a Bachelor of Science in mathematics in February of 1993 from Kyonggi University. In February 1995, he received a Master of Arts in statistics from ChungAng University. He moved to America to pursue his Ph.D. in August of 1997 in department of statistics at Texas A&M University. He received his Ph.D. in statistics under the direction of Dr. Jeffrey D. Hart and Dr. Marina Vannucci in August 2003. His permanent address is

990-105 DaeRim 2 Dong YoungDungPooGU

Seoul, Korea, 150-720