บทที่ 4

ผลการวิจัย

ผลการวิจัยครั้งนี้ เป็นผลจากการศึกษาเปรียบเทียบค่าประมาณของตัวแปรตาม ในสมการถดถอยเชิงเส้นอย่างง่าย โดยประมาณค่าพารามิเตอร์ด้วยวิธีการ 3 วิธี คือ วิธีกำลังสองต่ำสุด วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีการของบัคเลย์และเจมส์ ซึ่งการประมาณค่าพารามิเตอร์ ดังกล่าว นำมาใช้วิเคราะห์กับข้อมูลที่มีการตัดค่าทางขวาของตัวแปรตาม ความสามารถในการประมาณค่าพารามิเตอร์ของแต่ละวิธีการนั้น กระทำโดยการศึกษาค่าความคลาดเคลื่อนจากการประมาณค่าของตัวแปรตาม เมื่อคำนึงถึงสัดส่วนของข้อมูลที่ถูกตัดทิ้ง ขนาดตัวอย่าง การแจกแจงของค่าสังเกตที่ถูกตัดทิ้ง เพื่อหาข้อสรุปว่าวิธีการใดจะให้ค่าประมาณตัวแปรตามต่ำสุดในแต่ละสถานการณ์ที่เกิดขึ้นในการทดลอง โดยจะนำเสนอผลการวิจัยในรูปตารางและกราฟ และเพื่อความสะดวกในการอธิบายจะขอใช้สัญลักษณ์ แทนความหมายต่าง ๆ ที่จะปรากฏในตารางและกราฟ ดังต่อไปนี้

LS หมายถึง วิธีกำลังสองต่ำสุด

MLE หมายถึง วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด

BJ หมายถึง วิธีการของบัคเลย์และเจมส์

NM หมายถึง ขนาดตัวอย่างทั้งหมดทั้งที่ไม่ถูกตัดทิ้งและถูกตัดทิ้ง

N หมายถึง จำนวนค่าสังเกตที่ไม่ถูกตัดทิ้ง

M หมายถึง จำนวนค่าสังเกตที่ถูกตัดทิ้ง

การแจกแจงของค่าที่ถูกตัดทิ้ง (Censoring Distribution) รายละเอียดดังนี้

U	หมายถึง	แบบสม่ำเสมอ
TN	หมายถึง	แบบปกติตัดปลายทางซ้าย
TW	หมายถึง	แบบไวบูลล์ตัดปลายทางซ้าย
LI	หมายถึง	แบบเชิงเส้น

P	หมายถึง	สัดส่วนของข้อมูลที่ถูกตัดทิ้ง
RMSE	หมายถึง	ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการประมาณค่าของตัวแปรตาม
*	หมายถึง	ค่า RMSE ของการประมาณที่มีค่าต่ำที่สุด

4.1 สรุปผลการศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ทั้ง 3 วิธี

การเปรียบเทียบความสามารถในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อมีการตัดค่าทางขวาของตัวแปรตาม ทั้ง 3 วิธี เกณฑ์ที่ใช้ในการพิจารณาเปรียบเทียบ คือ ค่าความคลาดเคลื่อนระหว่างค่าประมาณของตัวแปรตาม กับค่าจริงในรูปของค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) วิธีการใดให้ค่า RMSE ต่ำกว่า จะเป็นวิธีการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่ายได้ดีกว่า ดังนั้น จึงสรุปผลได้ดังตารางที่ 4.1 - 4.5 และกราฟรูปที่ 4.1 - 4.5 ซึ่งแสดงค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการประมาณตัวแปรตามทั้ง 3 วิธี เมื่อสัดส่วนของข้อมูลที่ถูกตัดทิ้ง การแจกแจงของค่าที่ถูกตัดทิ้ง และขนาดตัวอย่างแตกต่างกัน

สำหรับขนาดตัวอย่างทุกขนาด (NM = 10, 15, 30, 50, 70) สัดส่วนของข้อมูลที่ถูกตัดทิ้งทุกระดับ (P = 10%, 20%, 30%, 40%) และการแจกแจงของค่าที่ถูกตัดทิ้งทุกแบบ คือ แบบสม่ำเสมอ แบบปกติตัดปลายทางซ้าย แบบไวบูลล์ตัดปลายทางซ้าย แบบเชิงเส้น (U, TN, TW, LI) สรุปผลได้ดังนี้

กรณีที่เป็นตัวอย่างขนาดเล็ก NM = 10, 15, 30 เมื่อสัดส่วนของข้อมูลที่ถูกตัดทิ้ง P = 10% และการแจกแจงของค่าที่ถูกตัดทิ้งทุกแบบ วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุดและวิธีการของบัคเลย์และเจมส์ ให้ค่า RMSE ใกล้เคียงกันมาก ส่วนวิธีกำลังสองต่ำสุดจะให้ค่า RMSE แตกต่างจากทั้งสองวิธีข้างต้นอย่างชัดเจน และเมื่อสัดส่วนของข้อมูลที่ถูกตัดทิ้งเพิ่มมากขึ้น ค่า RMSE จะมีค่ามากขึ้น และค่า RMSE ของแต่ละวิธีการจะมีค่าที่แตกต่างกันมากขึ้น ทั้งนี้ วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด ให้ค่า RMSE ต่ำกว่าวิธีการอื่น ๆ ในทุกสถานการณ์เมื่อเป็นกรณีตัวอย่างขนาดเล็ก วิธีการของบัคเลย์และเจมส์ ให้ค่า RMSE สูงกว่าวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด แต่ให้ค่า RMSE ต่ำกว่าวิธีกำลังสองต่ำสุด ในทุกสถานการณ์เมื่อเป็นกรณีตัวอย่างขนาดเล็ก และวิธีกำลังสองต่ำสุดให้ค่า RMSE สูงกว่าวิธีการอื่น ๆ ในทุกสถานการณ์เมื่อเป็นกรณีตัวอย่างขนาดเล็ก

กรณีที่เป็นตัวอย่างขนาดใหญ่ NM=50, 70 เมื่อสัดส่วนของข้อมูลที่ถูกตัดทิ้ง P=10%, 20%, 30%, 40% และการแจกแจงของค่าที่ถูกตัดทิ้งทุกแบบ วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีการของบัคเลย์และเจมส์ จะให้ค่า RMSE ที่ใกล้เคียงกันมาก ส่วนวิธีกำลังสองต่ำสุด จะให้ค่า RMSE แตกต่างจากทั้งสองวิธีข้างต้นอย่างชัดเจน และเมื่อสัดส่วนของข้อมูลที่ถูกตัดทิ้งเพิ่มมากขึ้น ค่า RMSE จะมีค่ามากขึ้น ทั้งนี้ วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุดให้ค่า RMSE ต่ำกว่าวิธีการอื่น ๆ ในทุกสถานการณ์เมื่อเป็นกรณีตัวอย่างขนาดใหญ่ วิธีการของบัคเลย์และเจมส์ ให้ค่า RMSE สูงกว่าวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด แต่ให้ค่า RMSE ต่ำกว่าวิธีกำลังสองต่ำสุด ในทุกสถานการณ์เมื่อเป็นกรณีตัวอย่างขนาดใหญ่ และวิธีกำลังสองต่ำสุดให้ค่า RMSE สูงกว่าวิธีการอื่น ๆ ในทุกสถานการณ์เมื่อเป็นกรณีตัวอย่างขนาดใหญ่ จึงสรุปผลแต่ละวิธีการดังนี้

4.2.1 วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด จะให้ค่าความคลาดเคลื่อน RMSE ของการประมาณค่าตัวแปรตามต่ำที่สุด ในทุก ๆ สถานการณ์ที่ศึกษา ดังนั้น วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุดจึงเป็นวิธีการที่มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ดีที่สุดในสามวิธีการที่ศึกษา

4.2.2 กรณีที่เป็นตัวอย่างขนาดเล็ก NM = 10, 15, 30 เมื่อสัดส่วนของข้อมูลที่ถูกตัดทิ้ง P = 10% และการแจกแจงของค่าที่ถูกตัดทิ้งทุกแบบ วิธีการของบัคเลย์และเจมส์จะให้ค่า RMSE ใกล้เคียงกับวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด กรณีที่เป็นตัวอย่างขนาดใหญ่ NM = 50, 70 เมื่อสัดส่วนของข้อมูลที่ถูกตัดทิ้ง P=10%, 20%, 30%, 40% และการแจกแจงของค่าที่ถูกตัดทิ้งทุกแบบ วิธีการของบัคเลย์และเจมส์จะให้ค่า RMSE ใกล้เคียงกับวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด ทั้งนี้วิธีการของบัคเลย์และเจมส์จะให้ค่าความคลาดเคลื่อน RMSE ของการประมาณค่าตัวแปรตามสูงกว่าวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุดในทุกสถานการณ์ที่ทำการศึกษา แต่วิธีการของบัคเลย์และเจมส์จะให้ค่าความคลาดเคลื่อน RMSE ของการประมาณค่าตัวแปรตามต่ำกว่าวิธีกำลังสองต่ำสุด ในทุกสถานการณ์ที่ศึกษา

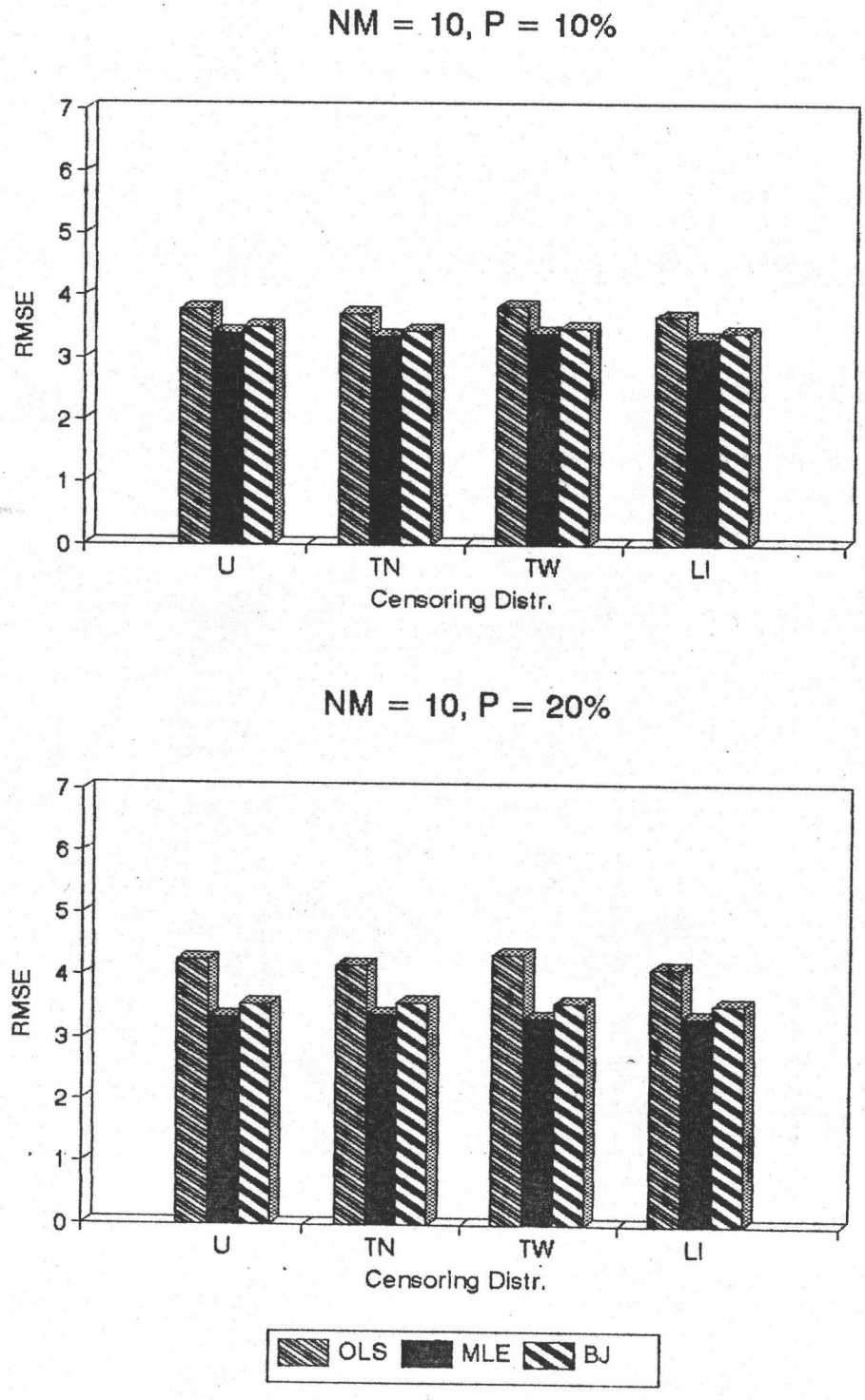
4.2.3 วิธีกำลังสองต่ำสุด จะให้ค่าความคลาดเคลื่อน RMSE ของการประมาณค่าตัวแปรตามสูงกว่าวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีการของบัคเลย์และเจมส์ ในทุกสถานการณ์ที่ทำการศึกษา

ตารางที่ 4.1 แสดงค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของการประมาณค่าที่ถูกตัดทิ้ง เมื่อขนาดตัวอย่างเท่ากับ 10 จำแนกตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P) และการแจกแจงของค่าที่ถูกตัดทิ้ง (Censoring Distribution)

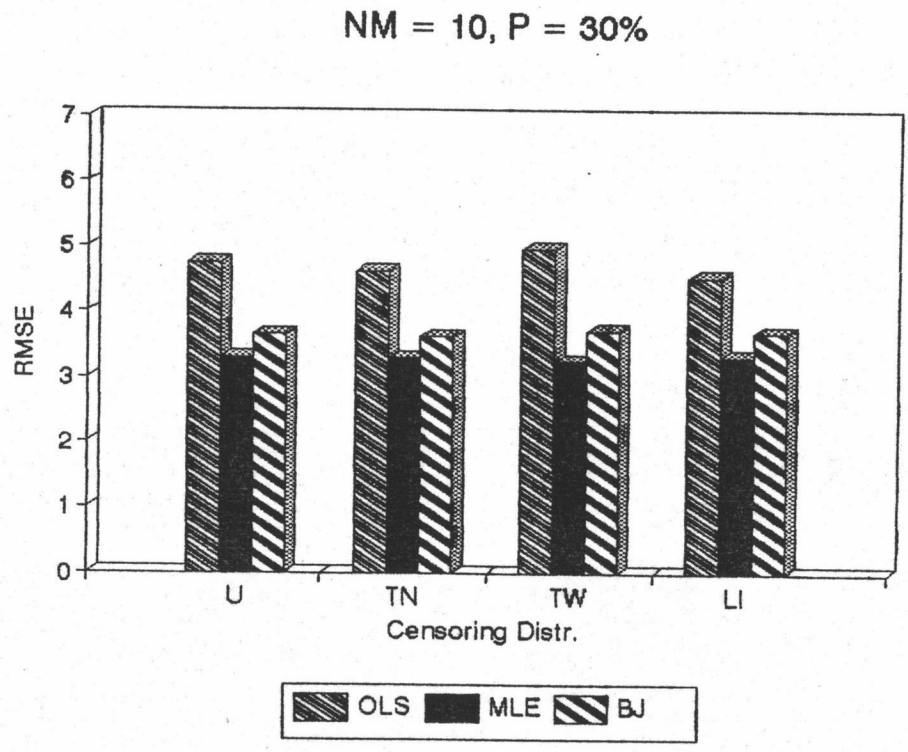
P	การแจกแจงของค่าที่ถูกตัดทิ้ง	RMSE		
		OLS	MLE	BJ
10%	U	3.787	3.407*	3.493
	TN	3.715	3.350*	3.430
	TW	3.836	3.404*	3.489
	LI	3.677	3.332*	3.416
20%	U	4.256	3.336*	3.546
	TN	4.185	3.384*	3.576
	TW	4.361	3.344*	3.582
	LI	4.147	3.363*	3.566
30%	U	4.741	3.318*	3.663
	TN	4.618	3.294*	3.631
	TW	4.963	3.254*	3.721
	LI	4.541	3.334*	3.696
40%	U	5.105	3.368*	3.641
	TN	4.847	3.233*	3.682
	TW	5.599	3.161*	3.786
	LI	4.582	3.233*	3.659

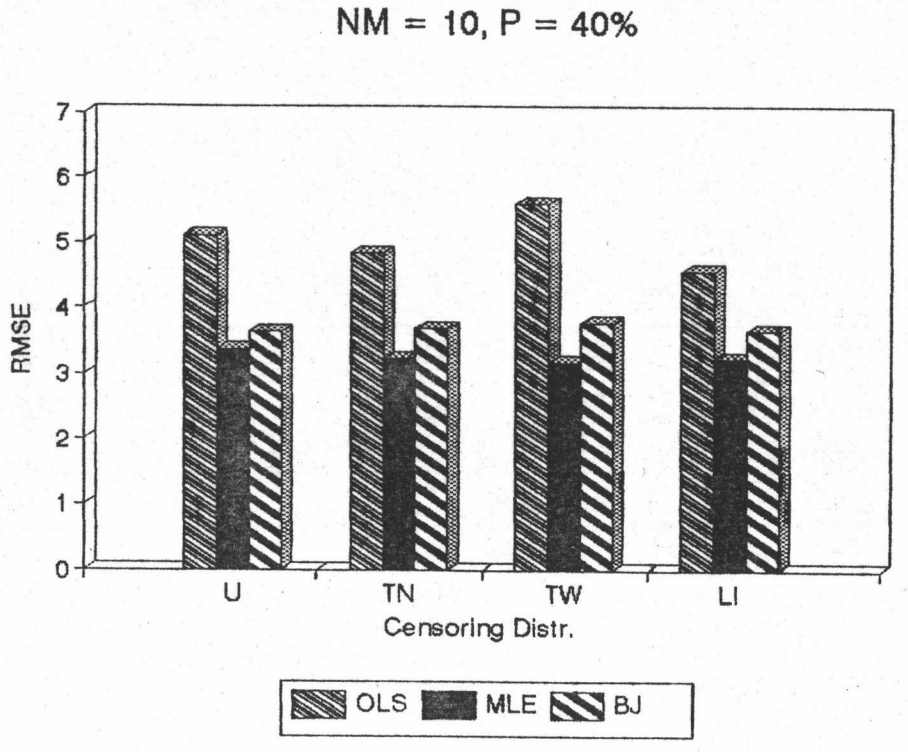
* หมายถึง ค่า RMSE ของการประมาณต่ำที่สุด

รูปที่ 4.1 การเปรียบเทียบ RMSE ของการประมาณ ทั้ง 3 วิธี ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าถูกตัดทิ้งทางขวา กรณีขนาดตัวอย่าง (NM) = 10 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง และสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P)



รูปที่ 4.1 (ต่อ)


NM = 10, P = 30%
RMSE
0
1
2
3
4
5
6
7
U
TN
TW
LI
Censoring Distr.
OLS
MLE
BJ

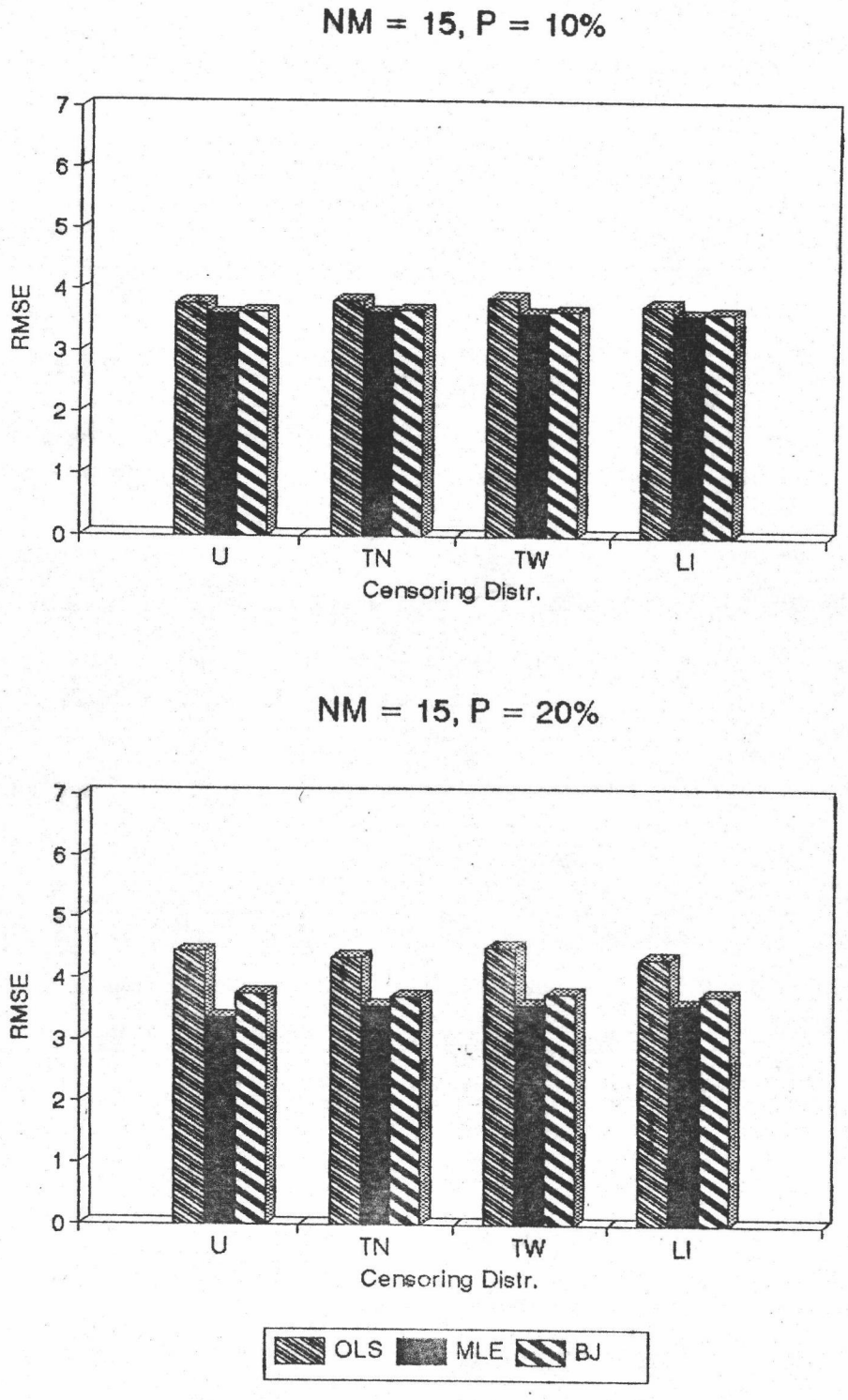

NM = 10, P = 40%
RMSE
0
1
2
3
4
5
6
7
U
TN
TW
LI
Censoring Distr.
OLS
MLE
BJ

ตารางที่ 4.2 แสดงค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของการประมาณค่าที่ถูกตัดทิ้ง เมื่อขนาดตัวอย่างเท่ากับ 15 จำแนกตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P) และการแจกแจงของค่าที่ถูกตัดทิ้ง

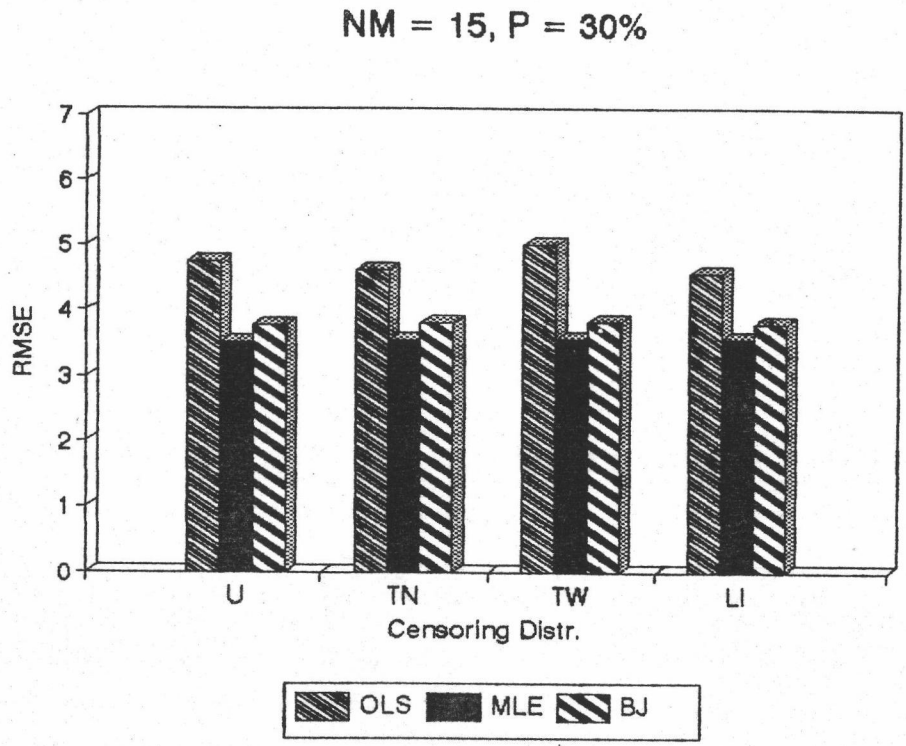
P	การแจกแจงของค่าที่ถูกตัดทิ้ง	RMSE		
		OLS	MLE	BJ
10%	U	3.791	3.604*	3.636
	TN	3.836	3.649*	3.678
	TW	3.872	3.626*	3.654
	LI	3.775	3.614*	3.644
20%	U	4.452	3.380*	3.772
	TN	4.348	3.562*	3.723
	TW	4.531	3.580*	3.770
	LI	4.342	3.601*	3.758
30%	U	4.751	3.513*	3.782
	TN	4.620	3.553*	3.816
	TW	5.012	3.564*	3.835
	LI	4.581	3.558*	3.795
40%	U	5.333	3.622*	3.870
	TN	5.093	3.592*	3.883
	TW	5.896	3.470*	4.064
	LI	4.994	3.579*	3.887

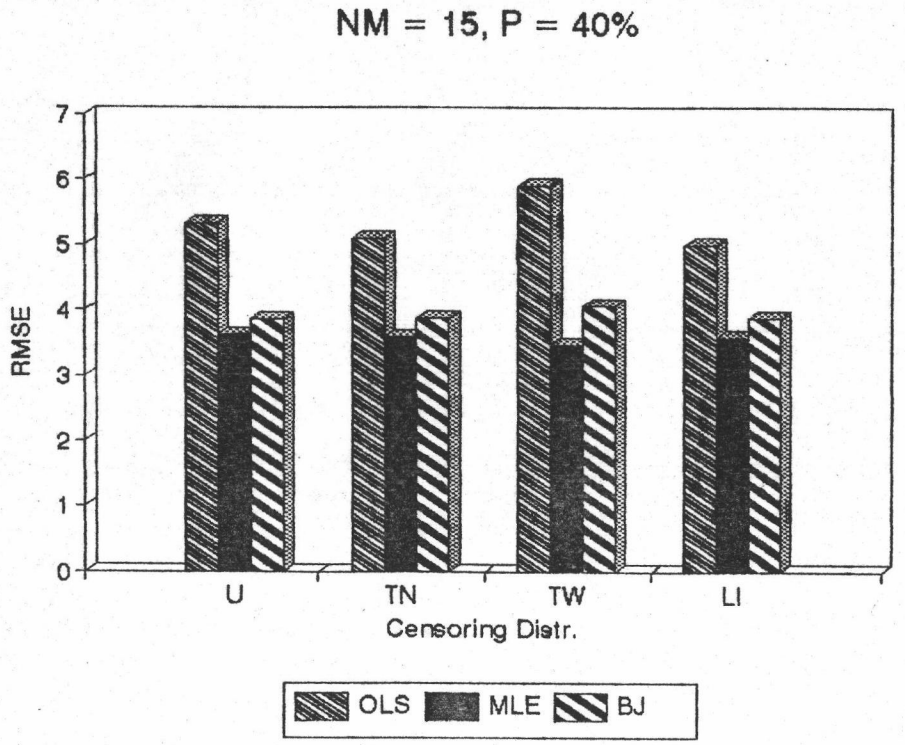
* หมายถึง ค่า RMSE ของการประมาณต่ำที่สุด

รูปที่ 4.2 การเปรียบเทียบ RMSE ของการประมาณ ทั้ง 3 วิธี ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าถูกตัดทิ้งทางขวา กรณีขนาดตัวอย่าง (NM) = 15 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง และสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P)



รูปที่ 4.2 (ต่อ)


NM = 15, P = 30%
RMSE
0
1
2
3
4
5
6
7
U
TN
TW
LI
Censoring Distr.
OLS
MLE
BJ

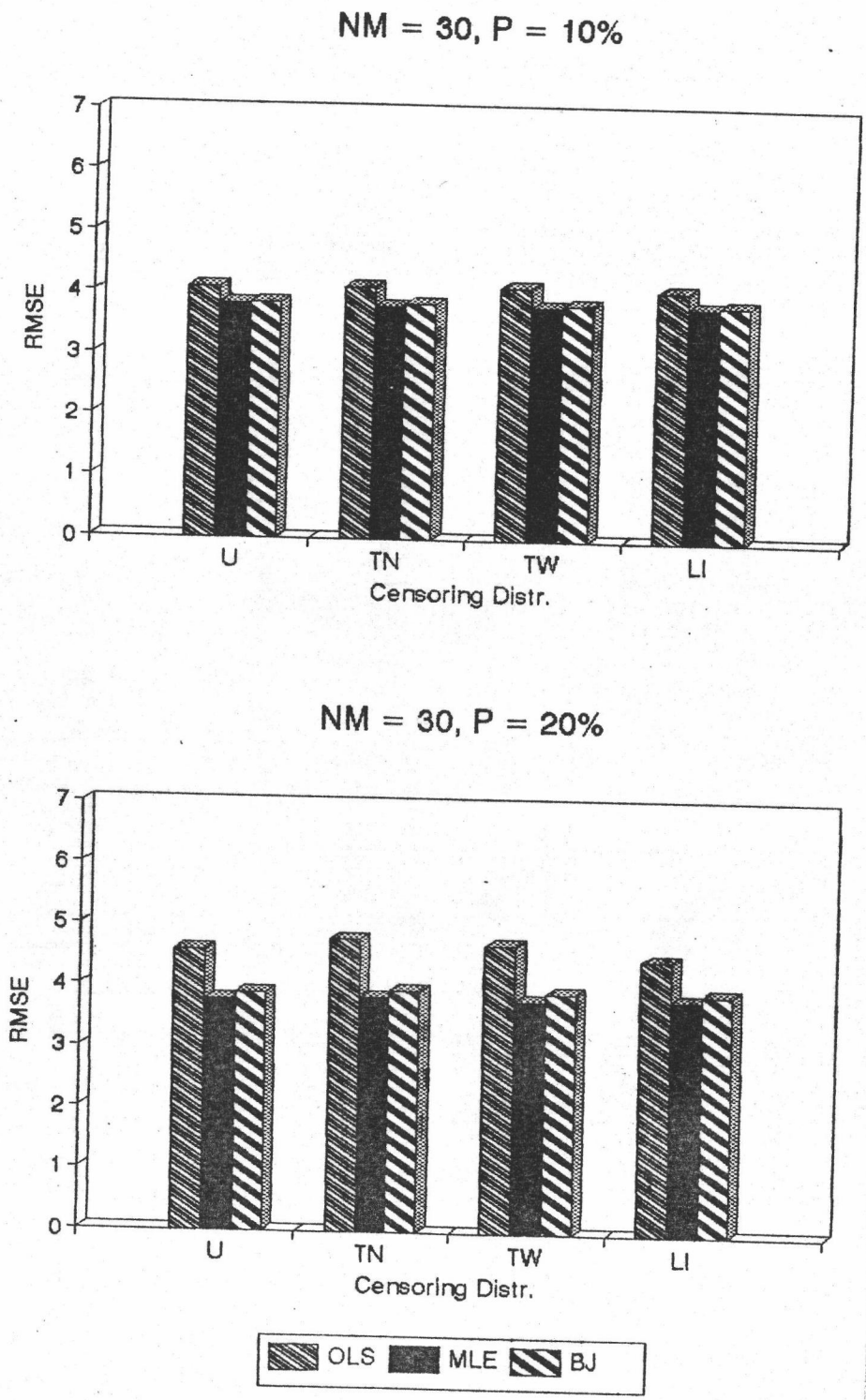

NM = 15, P = 40%
RMSE
0
1
2
3
4
5
6
7
U
TN
TW
LI
Censoring Distr.
OLS
MLE
BJ

ตารางที่ 4.3 แสดงค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของการประมาณค่าที่ถูกตัดทิ้ง เมื่อขนาดตัวอย่างเท่ากับ 30 จำแนกตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P) และการแจกแจงของค่าที่ถูกตัดทิ้ง

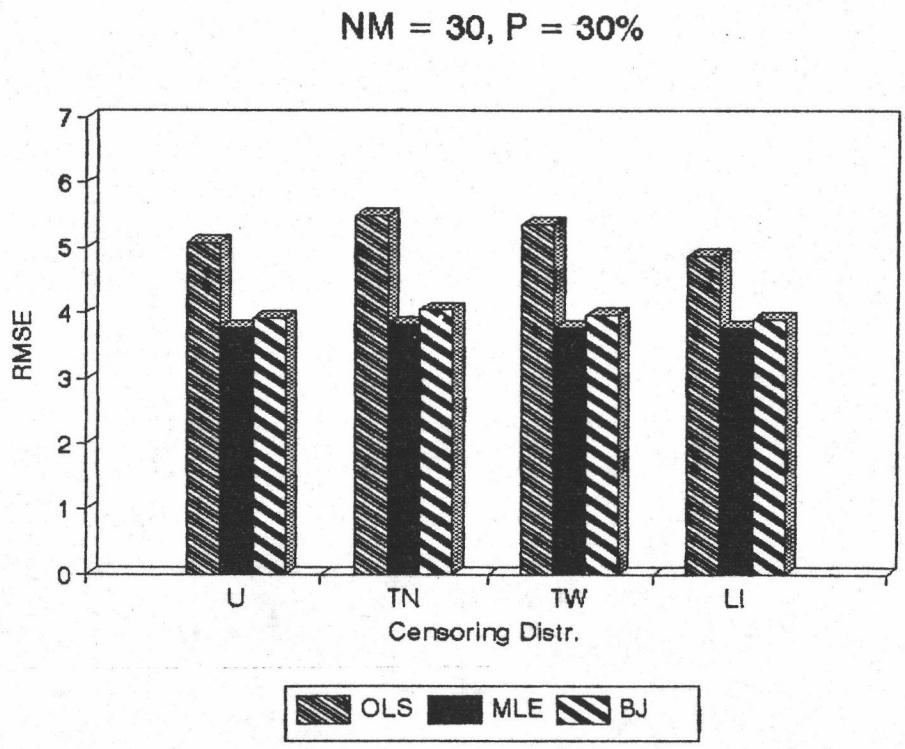
P	การแจกแจงของค่าที่ถูกตัดทิ้ง	RMSE		
		OLS	MLE	BJ
10%	U	4.101	3.826*	3.854
	TN	4.116	3.808*	3.840
	TW	4.127	3.822*	3.852
	LI	4.096	3.836*	3.861
20%	U	4.590	3.782*	3.896
	TN	4.775	3.825*	3.937
	TW	4.694	3.808*	3.928
	LI	4.474	3.827*	3.928
30%	U	5.076	3.790*	3.934
	TN	5.487	3.830*	4.060
	TW	5.357	3.783*	3.981
	LI	4.891	3.782*	3.914
40%	U	5.523	3.795*	3.960
	TN	6.154	3.775*	4.055
	TW	6.095	3.787*	4.061
	LI	5.179	3.772*	3.893

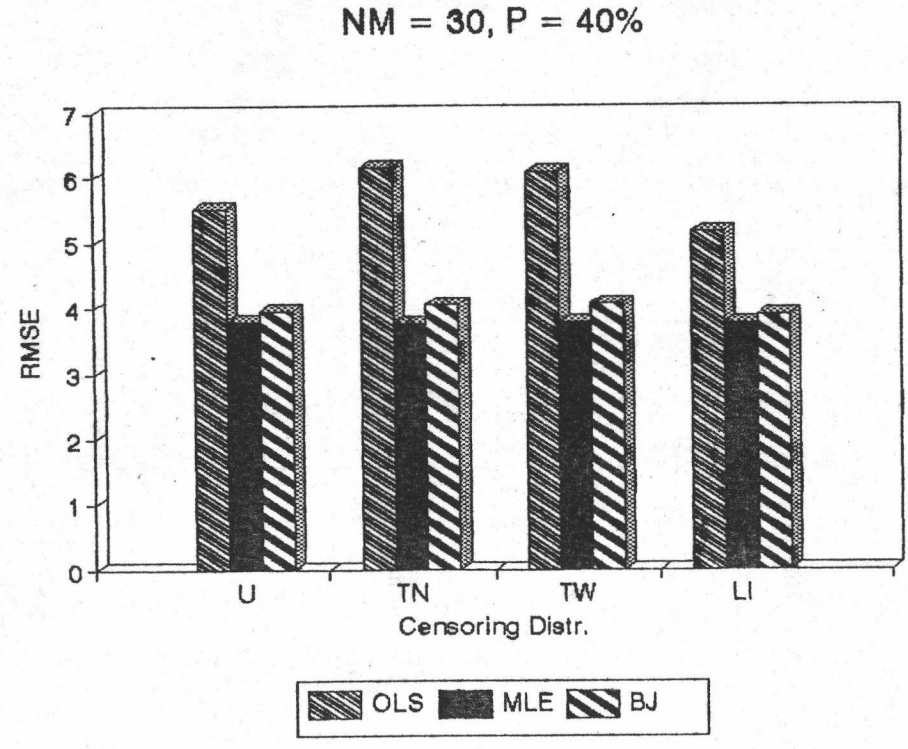
* หมายถึง ค่า RMSE ของการประมาณต่ำที่สุด

รูปที่ 4.3 การเปรียบเทียบ RMSE ของการประมาณ ทั้ง 3 วิธี ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าถูกตัดทิ้งทางขวา กรณีขนาดตัวอย่าง (NM) = 30 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง และสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P)



รูปที่ 4.3 (ต่อ)


NM = 30, P = 30%
RMSE
0
1
2
3
4
5
6
7
U
TN
TW
LI
Censoring Distr.
OLS
MLE
BJ

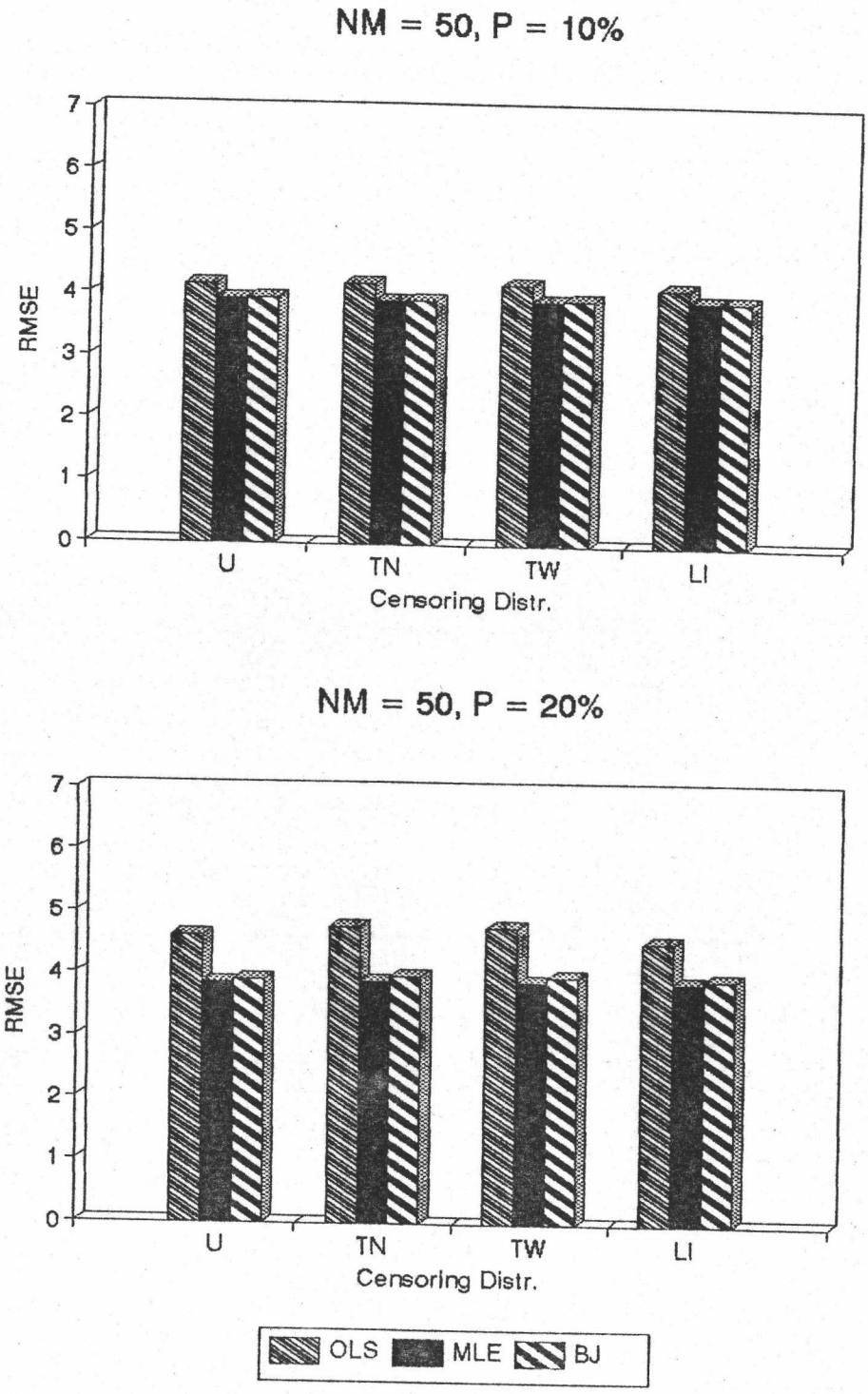

NM = 30, P = 40%
RMSE
0
1
2
3
4
5
6
7
U
TN
TW
LI
Censoring Distr.
OLS
MLE
BJ

ตารางที่ 4.4 แสดงค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของการประมาณค่าที่ถูกตัดทิ้ง เมื่อขนาดตัวอย่างเท่ากับ 50 จำแนกตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P) และการแจกแจงของค่าที่ถูกตัดทิ้ง

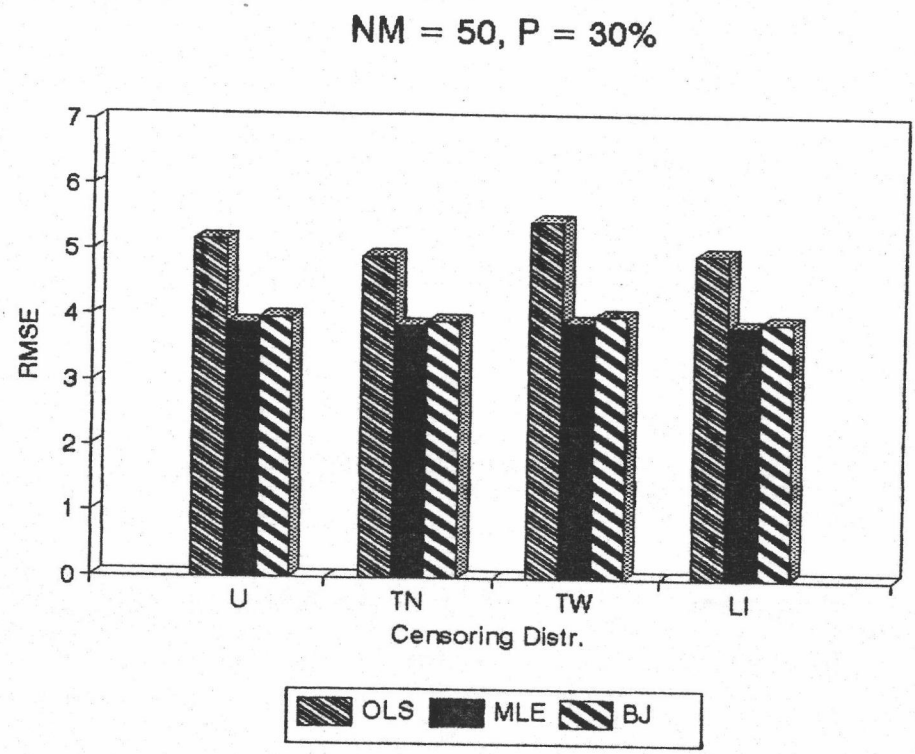
P	การแจกแจงของค่าที่ถูกตัดทิ้ง	RMSE		
		OLS	MLE	BJ
10%	U	4.171	3.923*	3.940
	TN	4.196	3.911*	3.930
	TW	4.197	3.923*	3.941
	LI	4.145	3.936*	3.950
20%	U	4.616	3.859*	3.916
	TN	4.770	3.895*	3.977
	TW	4.760	3.897*	3.975
	LI	4.546	3.896*	3.941
30%	U	5.168	3.884*	3.979
	TN	4.917	3.858*	3.933
	TW	5.462	3.903*	4.021
	LI	4.971	3.866*	3.925
40%	U	5.617	3.874*	3.963
	TN	6.284	3.931*	4.075
	TW	6.159	3.846*	3.962
	LI	5.257	3.851*	3.879

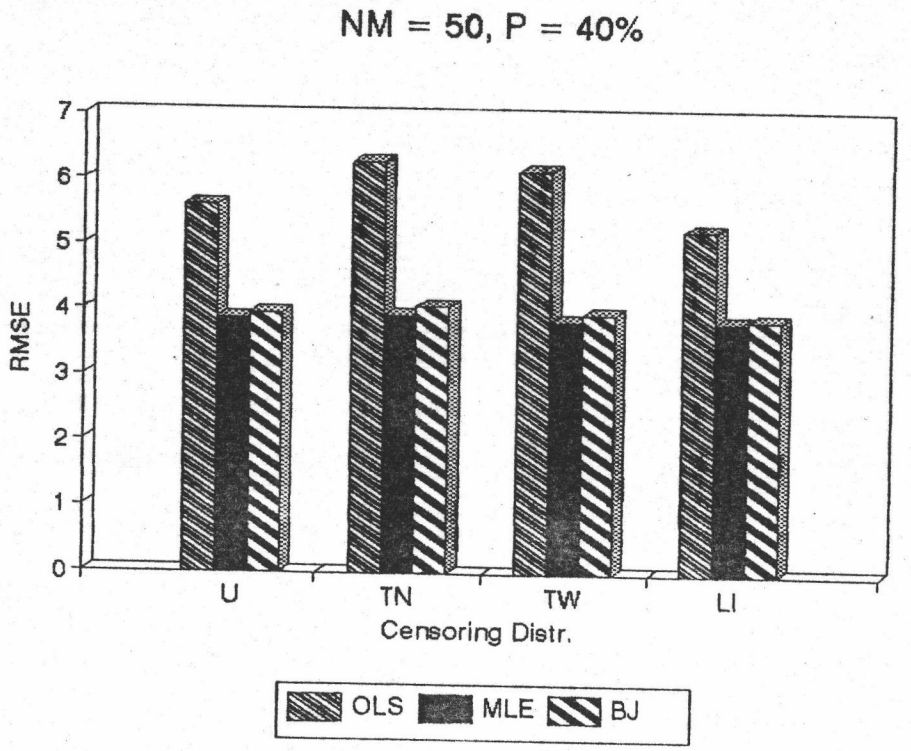
* หมายถึง ค่า RMSE ของการประมาณต่ำที่สุด

รูปที่ 4.4 การเปรียบเทียบ RMSE ของการประมาณ ทั้ง 3 วิธี ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าถูกตัดทิ้งทางขวา กรณีขนาดตัวอย่าง (NM) = 50 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง และสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P)



รูปที่ 4.4 (ต่อ)


NM = 50, P = 30%
RMSE
0
1
2
3
4
5
6
7
U
TN
TW
LI
Censoring Distr.
OLS
MLE
BJ

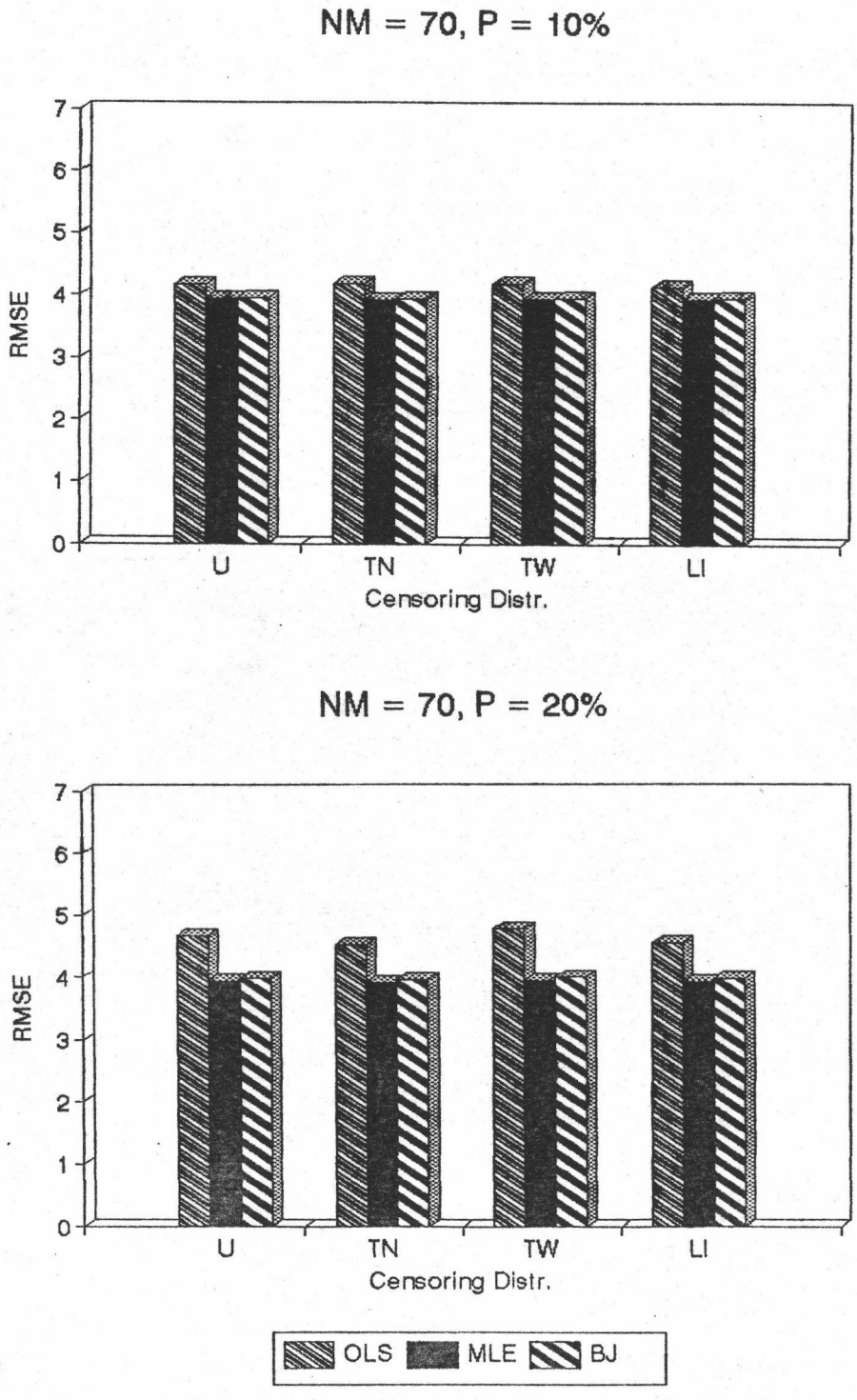

NM = 50, P = 40%
RMSE
0
1
2
3
4
5
6
7
U
TN
TW
LI
Censoring Distr.
OLS
MLE
BJ

ตารางที่ 4.5 แสดงค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของการประมาณค่าที่ถูกตัดทิ้ง เมื่อขนาดตัวอย่างเท่ากับ 70 จำแนกตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P) และการแจกแจงของค่าที่ถูกตัดทิ้ง

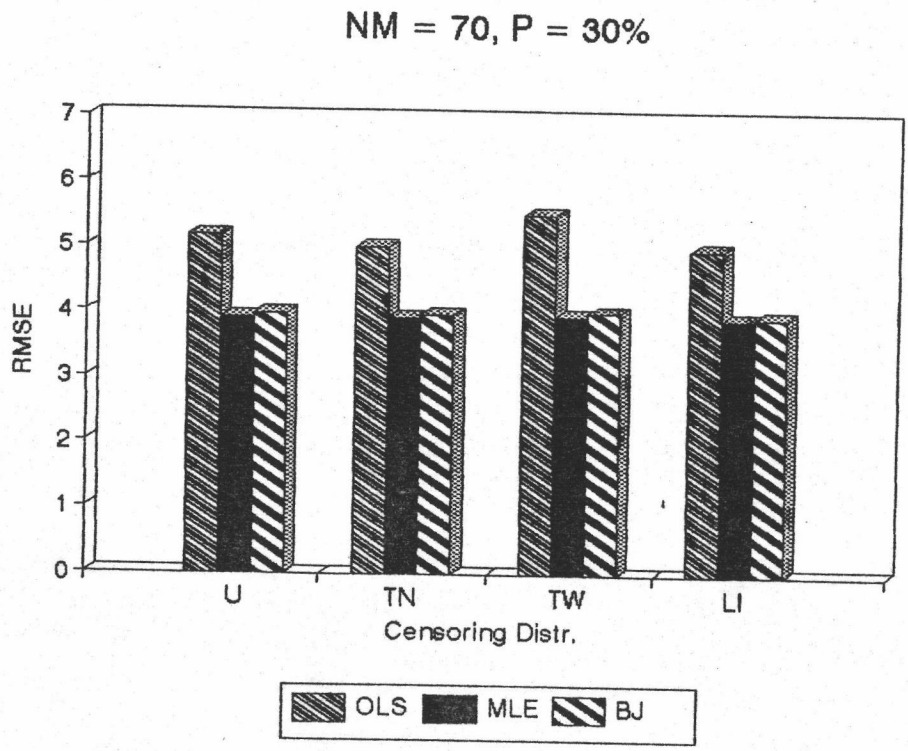
P	การแจกแจงของค่าที่ถูกตัดทิ้ง	RMSE		
		OLS	MLE	BJ
10%	U	4.175	3.945*	3.959
	TN	4.202	3.932*	3.947
	TW	4.197	3.943*	3.958
	LI	4.151	3.949*	3.962
20%	U	4.664	3.939*	3.982
	TN	4.533	3.911*	3.954
	TW	4.778	3.949*	3.998
	LI	4.553	3.931*	3.971
30%	U	5.180	3.904*	3.975
	TN	4.997	3.911*	3.962
	TW	5.507	3.935*	4.012
	LI	4.980	3.904*	3.950
40%	U	5.613	3.879*	3.928
	TN	5.320	3.885*	3.900
	TW	6.202	3.938*	4.015
	LI	5.278	3.888*	3.920

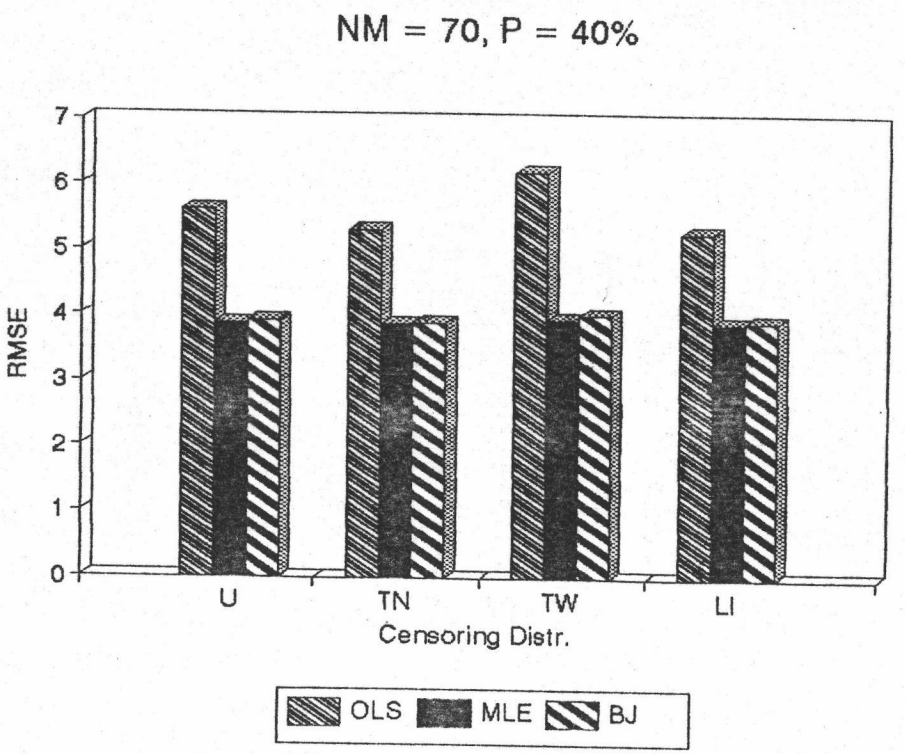
* หมายถึง ค่า RMSE ของการประมาณต่ำที่สุด

รูปที่ 4.5 การเปรียบเทียบ RMSE ของการประมาณ ทั้ง 3 วิธี ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าถูกตัดทิ้งทางขวา กรณีขนาดตัวอย่าง (NM) = 70 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง และสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P)



รูปที่ 4.5 (ต่อ)


NM = 70, P = 30%
RMSE
0
1
2
3
4
5
6
7
U
TN
TW
LI
Censoring Distr.
OLS
MLE
BJ


NM = 70, P = 40%
RMSE
0
1
2
3
4
5
6
7
U
TN
TW
LI
Censoring Distr.
OLS
MLE
BJ