

ปฏิบัติการที่ 8

การใช้เครื่องมือวิเคราะห์ข้อมูลอนุกรมเวลา

วัตถุประสงค์

1. เพื่อให้สามารถใช้เครื่องมือวิเคราะห์ข้อมูลอนุกรมเวลาได้
2. เพื่อให้รู้วิธีการเพิ่มหรือพัฒนาเครื่องมือการวิเคราะห์ข้อมูลด้วยการพัฒนาคำสั่งขนาดเล็ก

ชุดข้อมูลปฏิบัติการ

- ชุดข้อมูล VIX Prices (สำหรับการสาธิตและการฝึกปฏิบัติการ)

ขั้นตอนปฏิบัติการ

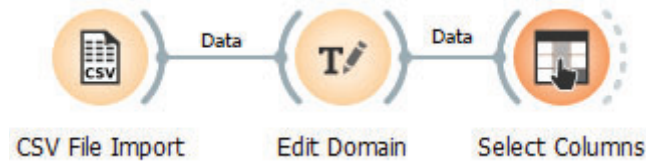
ขั้นตอนปฏิบัติการ มีดังนี้

1. เปิดโปรแกรม Orange
2. ทำการบันทึก workspace โดยไปที่เมนู File เลือก Save จากนั้นทำการตั้งชื่อไฟล์ในรูปแบบ Practice_08_id.ows โดยแทน id ด้วยรหัสนักศึกษา แล้วกดปุ่ม Save
3. นำชุดข้อมูลจากแฟ้มข้อมูล vix_prices.csv เข้าสู่โปรแกรม Orange โดยใช้โมดูล CSV File Import
4. ตรวจสอบและเปลี่ยนชนิดข้อมูลของตัวแปร โดยใช้โมดูล Edit Domain (ดูปฏิบัติการที่ 2) กำหนดให้แต่ละตัวแปรต้องมีชนิดข้อมูล ดังนี้

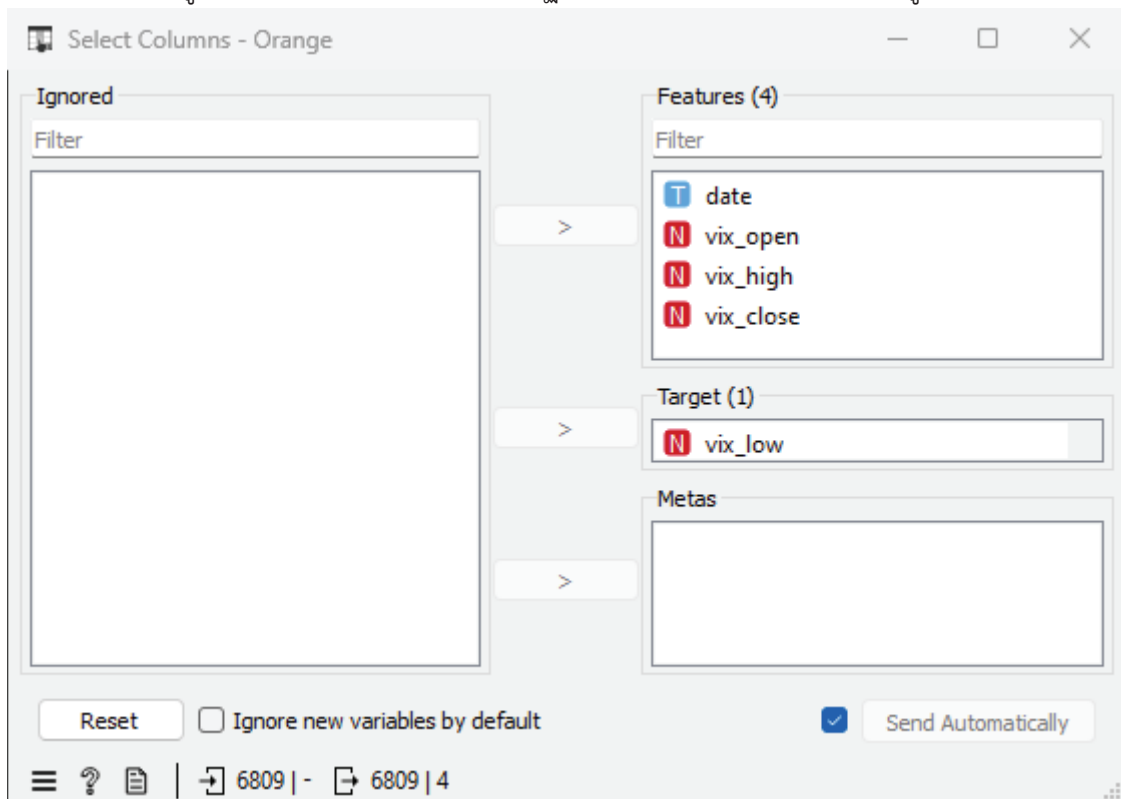
| ตัวแปร | ชนิดข้อมูล |
|-----------|-----------------|
| date | Time |
| vix_open | Numeric Feature |
| vix_high | Numeric Feature |
| vix_low | Numeric Feature |
| vix_close | Numeric Feature |

1. ในปฏิบัติการนี้จะสร้างโมเดล ARIMA ในการวิเคราะห์ข้อมูลอนุกรมเวลาเพื่อการทำนายค่าในอนาคต โดยทำการวิเคราะห์ข้อมูลราคาต่ำสุดของดัชนีตลาดซื้อขายอนุพันธ์ Chicago Board Options Exchange (CBOE)

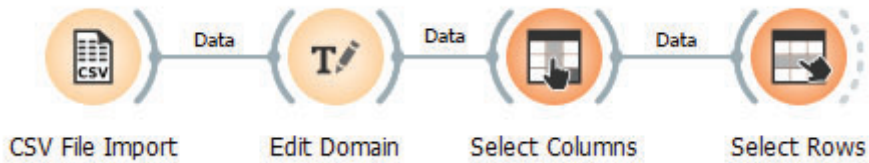
2. ใช้โมดูล Select Columns ในการเลือกตัวแปรต้นและตัวแปรตาม โดยคลิกเลือกโมดูล Select Columns จะปรากฏโมดูลใน workspace จากนั้นคลิกเชื่อมโมดูล Edit Domain จากด้าน output เข้าสู่โมดูล Select Columns ด้าน input ดังรูป



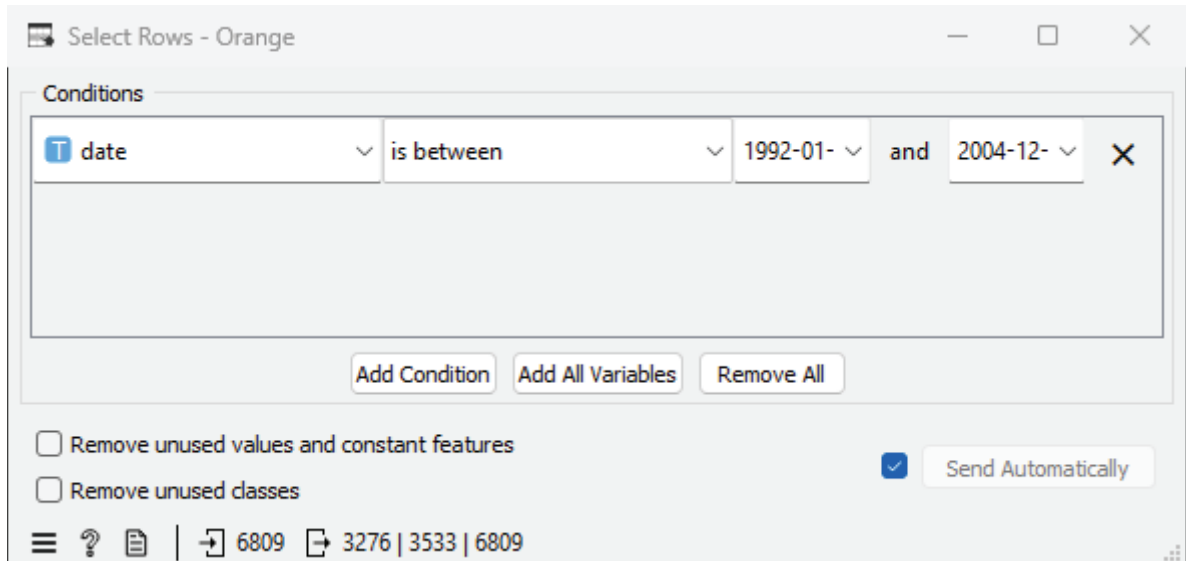
3. ดับเบิลคลิกที่โมดูล Select Columns จากปรากฏหน้าต่างสำหรับเลือกตัวแปร ดังรูป



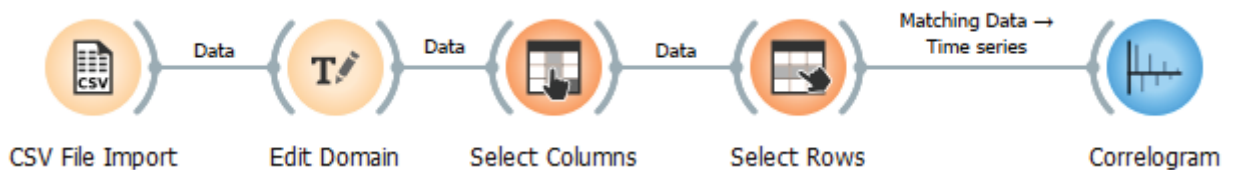
4. ให้ข้อมูลในชุดข้อมูลเรียนรู้เป็นข้อมูลหลังปี ค.ศ.2005 ส่วนข้อมูลในชุดข้อมูลทดสอบเป็นข้อมูลระหว่างปี ค.ศ.2005 ถึง 2019 ทำได้โดยใช้ Select Rows ในการเลือกแถวข้อมูล โดยคลิกเลือกโมดูล Select Rows จะปรากฏโมดูลใน workspace จากนั้นคลิกเชื่อมโมดูล Edit Domain (นำผลลัพธ์จากโมดูล Edit Domain ไปใช้ต่อ) จากด้าน output เข้าสู่โมดูล Select Rows ด้าน input ดังรูป



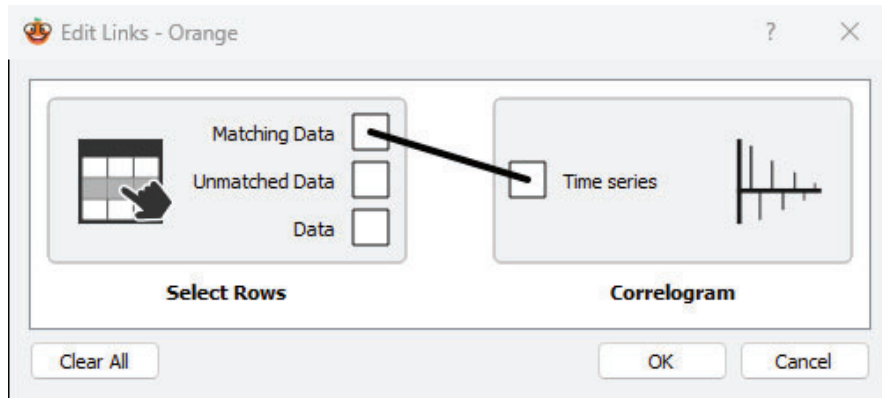
5. จากนั้น ดับเบิลคลิกที่เส้นเชื่อมระหว่างโมดูล Select Rows จากปรากฏหน้าต่างดังรูปด้านล่าง กำหนดเงื่อนไขของการเลือกแถวข้อมูล บนตัวแปร Date โดยมีเงื่อนไขเป็นช่วงวันที่ระหว่าง (is between) 1992-01-01 และ 2004-12-31



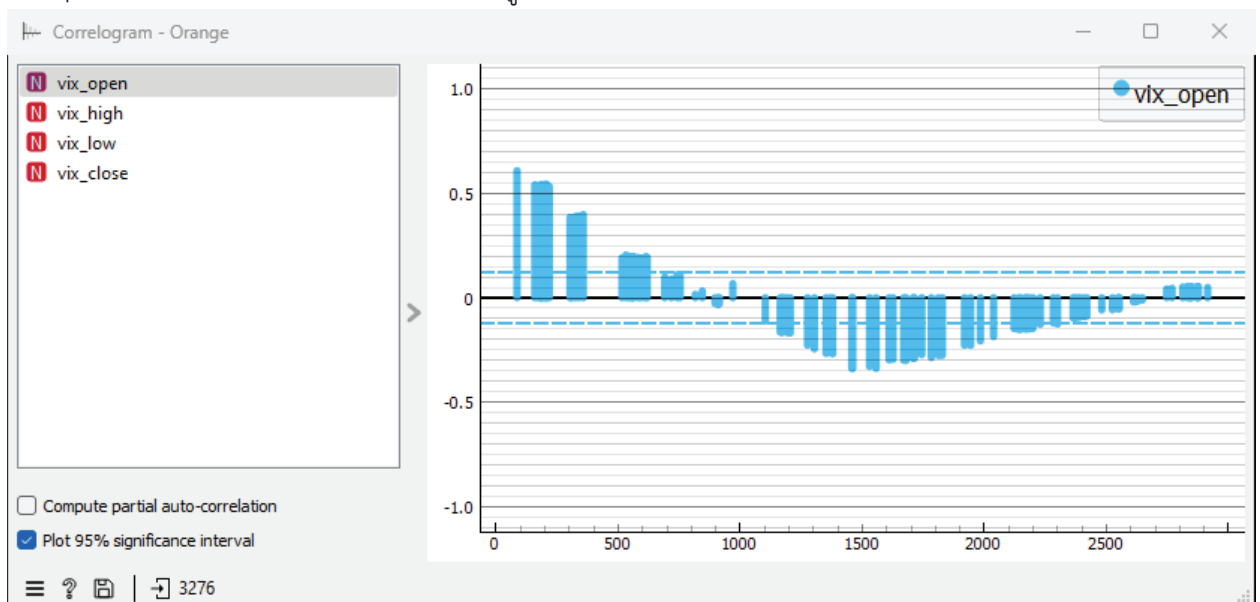
6. ต่อมาทำการคำนวณค่า autocorrelation function (ACF) หรือ partial autocorrelation function (PACF) โดยใช้โมดูล Correlogram คลิกเลือกโมดูล Correlogram จะปรากฏโมดูลใน workspace จากนั้นคลิกเชื่อมโมดูล Select Rows จากด้าน output เข้าสู่โมดูล Correlogram ด้าน input ดังรูป



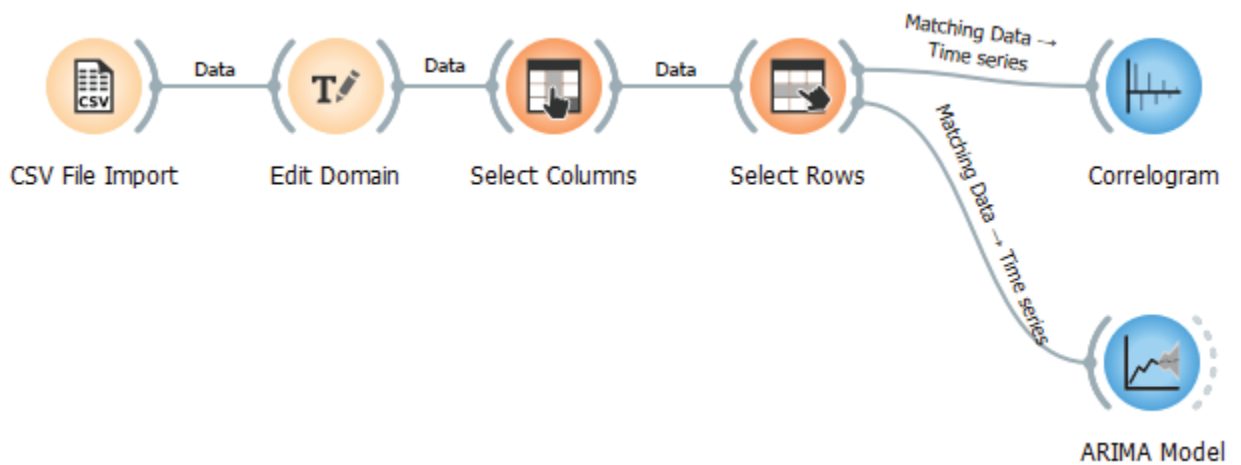
7. จากนั้น ดับเบิลคลิกที่เส้นเชื่อมระหว่างโมดูล Select Rows และ Correlogram จากปรากฏหน้าต่างดังรูปด้านล่าง ให้ลากเส้นเชื่อมระหว่าง Matching Data และ Time series เป็นการกำหนดให้ข้อมูลระหว่าง 1992-01-01 และ 2004-12-31 มาใช้ในการคำนวณค่า autocorrelation function หรือ partial autocorrelation function



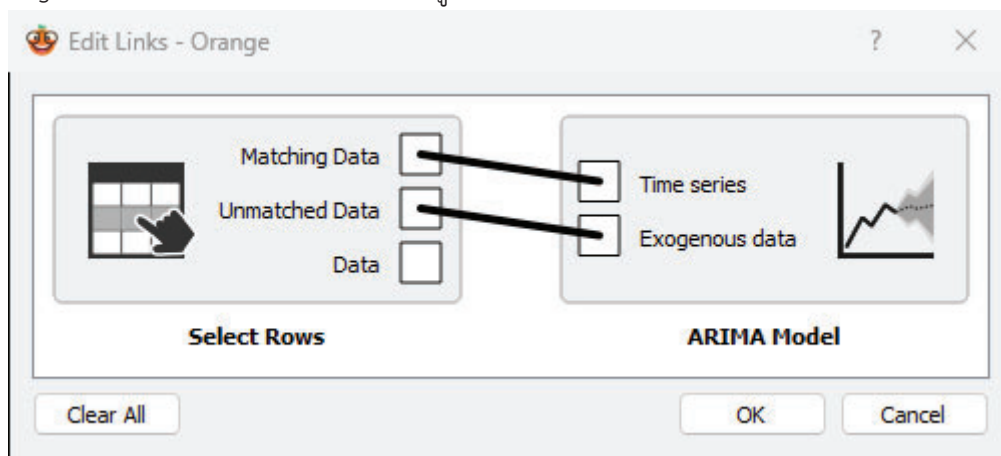
8. ดับเบิลคลิกโมดูล Correlogram จะปรากฏหน้าต่างแสดงผลการคำนวณค่า autocorrelation function หรือ partial autocorrelation function ดังรูป



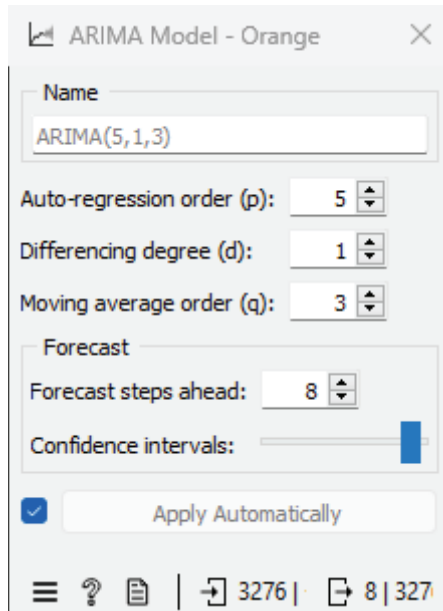
9. จากหน้าต่างนี้ สามารถเลือกตัวแปรที่ต้องการศึกษาได้จากรายการด้านซ้าย และหากต้องการคำนวณค่า partial autocorrelation สามารถทำได้โดยคลิกเลือก Compute partial auto-correlation
10. ต่อมาทำการสร้างโมเดล ARIMA สำหรับการวิเคราะห์ข้อมูลอนุกรมเวลา โดยใช้โมดูล ARIMA Model คลิกเลือกโมดูล ARIMA Model จะปรากฏโมดูลใน workspace จากนั้นคลิกเชื่อมโมดูล Select Rows จากด้าน output เข้าสู่โมดูล ARIMA Model ด้าน input ดังรูป



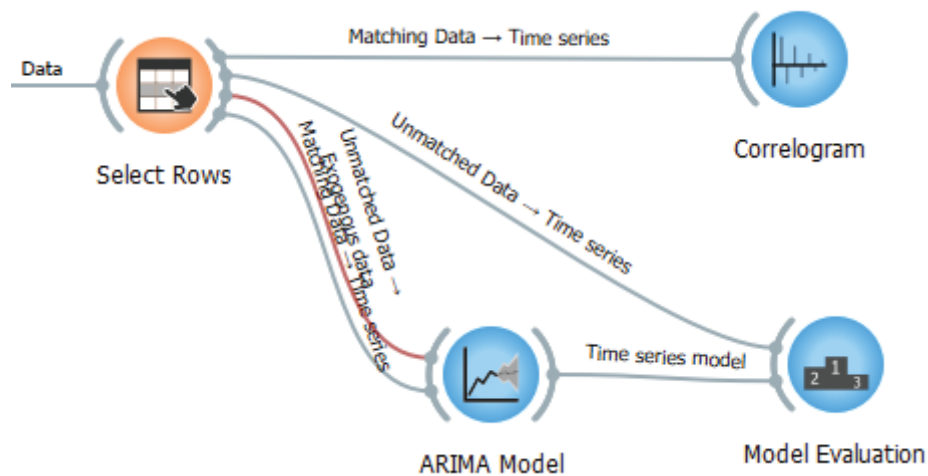
11. จากนั้น ดับเบิลคลิกที่เส้นเชื่อมระหว่างโมดูล Select Rows และ ARIMA Model จากปรากฏหน้าต่างดังรูปด้านล่าง ให้ลากเส้นเชื่อมระหว่าง Matching Data และ Time series เพื่อกำหนดให้นำข้อมูลระหว่าง 1992-01-01 และ 2004-12-31 มาใช้ในสอนแบบจำลอง และลากเส้นเชื่อมระหว่าง Unmatched Data และ Exogenous data เพื่อกำหนดให้นำข้อมูลตั้งแต่ 2005-01-01 มาใช้ในการทำนาย



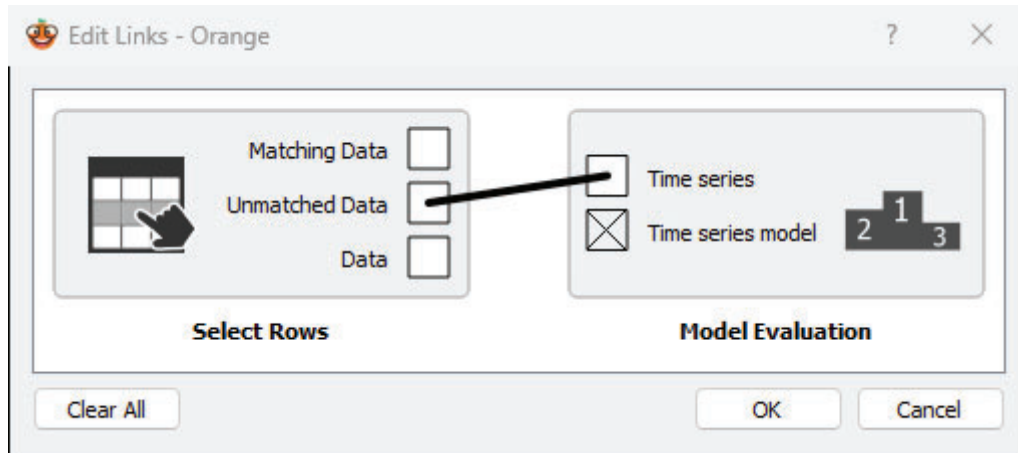
12. ดับเบิลคลิกโมดูล ARIMA Model ทำการกำหนดค่าพารามิเตอร์ของแบบจำลอง ดังรูป



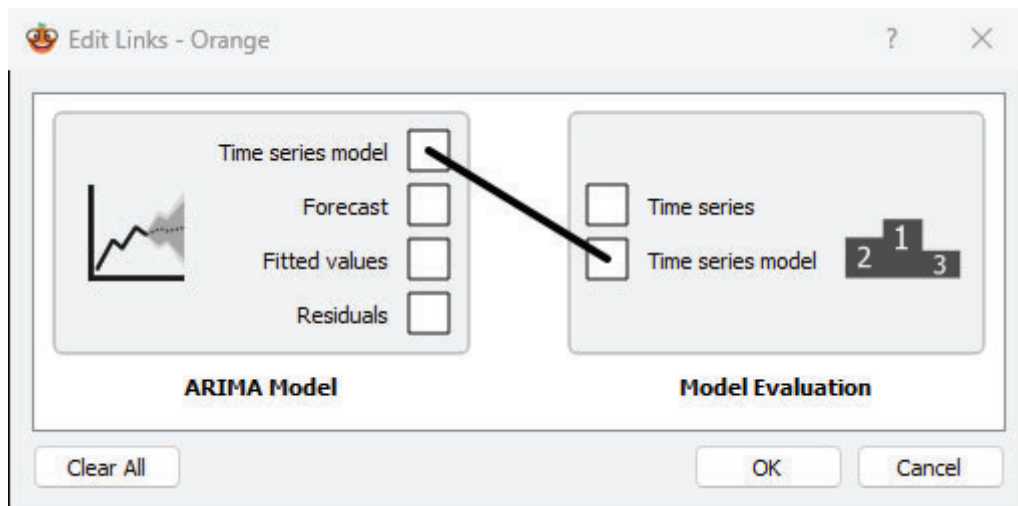
13. ทำการประเมินประสิทธิภาพของแบบจำลอง โดยใช้โมดูล Model Evaluation คลิกเลือกโมดูล Model Evaluation จะปรากฏโมดูลใน workspace จากนั้นคลิกเชื่อมโมดูล ARIMA Model จากด้าน output เข้าสู่โมดูล Model Evaluation ด้าน input และเชื่อมโมดูล Select Rows จากด้าน output เข้าสู่โมดูล Model Evaluation ด้าน input ดังรูป



14. จากนั้น ดับเบิลคลิกที่เส้นเชื่อมระหว่างโมดูล Select Rows และ Model Evaluation จากปรากฏหน้าต่างดังรูปด้านล่าง ให้ลากเส้นเชื่อมระหว่าง Unmatched Data และ Time series เพื่อกำหนดให้นำข้อมูลระหว่างตั้งแต่ 2005-01-01 มาใช้ในการทดสอบประสิทธิภาพของแบบจำลอง



15. จากนั้น ดับเบิลคลิกที่เส้นเชื่อมระหว่างโมเดล ARIMA Model และ Model Evaluation จากปรากฏหน้าต่างดังรูปด้านล่าง ให้ลากเส้นเชื่อมระหว่าง Time series model และ Time series model เพื่อกำหนดให้นำแบบจำลอง ARIMA มาทดสอบประสิทธิภาพ



16. ดับเบิลคลิกโมเดล Model Evaluation จะปรากฏผลการประเมินประสิทธิภาพ ดังรูป

| | RMSE | MAE | MAPE | POCID | R ² | AIC | BIC |
|--------------------------|-------|-------|-------|-------|----------------|-------|-------|
| ARIMA(5,1,3) | 2.838 | 2.440 | 0.108 | 50.0 | 0.391 | 11898 | 11953 |
| ARIMA(5,1,3) (in-sample) | 1.334 | 0.542 | 0.047 | 51.1 | 0.975 | 11980 | 12036 |

แบบฝึกปฏิบัติการ

ให้นักศึกษาทำแบบฝึกปฏิบัติการ ตามลำดับขั้นตอนต่อไปนี้

1. กำหนดให้นักศึกษาทำแบบฝึกปฏิบัติการนี้ ต่อจากการทดลองสาธิต โดยให้นักศึกษาใช้ชุดข้อมูล VIX Prices สร้างโมเดล ARIMA พร้อมทั้งคำนวณและแสดงค่า ACF และ PACF สำหรับการวิเคราะห์อนุกรมเวลา บนตัวแปร ราคาสูงสุดของ vix (vix_high) เพื่อทำนายค่าราคาสูงสุดของ VIX ตั้งแต่วันที่ 1 มกราคม ค.ศ.2005 ถึง วันที่ 15 มกราคม ค.ศ. 2019
2. คำนวณค่า ACF และ PACF
3. สร้างแบบจำลอง ARIMA(5,1,5)
4. ทำการทดสอบประสิทธิภาพของแบบจำลอง ARIMA
5. ทดลองเปลี่ยนค่าไฮเปอร์พารามิเตอร์แบบจำลอง ARIMA แล้วสังเกตผลการวัดประสิทธิภาพ
6. **สิ่งที่ต้องส่งเป็นการบ้าน** ทำการบันทึกไฟล์ workspace ของนักศึกษา โดยตั้งชื่อไฟล์ในรูปแบบ Lab_08_id.ows โดยแทน id ด้วยรหัสนักศึกษา ส่งผ่านเว็บไซต์ <http://hw.cs.science.cmu.ac.th>