

เศรษฐกิจ' ธรรมศาสตร์ ตลาดวิชา

โปรแกรมอาร์ (R) ซอฟต์แวร์โอเพนซอร์สสำหรับการวิเคราะห์ข้อมูล

อ.ดร.เฉลิมพงษ์ คงเจริญ

ผมเริ่มเรียนวิชาเศรษฐมิติเมื่อสิบห้าปีก่อน ขณะนั้นโปรแกรมคอมพิวเตอร์ที่ใช้ประมาณค่าสัมประสิทธิ์ต่างๆ ต้องใช้โปรแกรมสำเร็จรูปที่มีอยู่ในห้องคอมพิวเตอร์ของคณะเศรษฐศาสตร์ หนึ่งในโปรแกรมหลักที่นักเศรษฐศาสตร์ไทยสมัยนั้นใช้คือ EViews (www.eviews.com) ซึ่งเป็นโปรแกรมที่ต้องได้รับอนุญาตหรือต้องใช้แบบผิดกฎหมาย ทางเลือกอื่นๆ เช่น การหาโปรแกรมแบบเดียวกันทางอินเทอร์เน็ตเหมือนสมัยนี้ทำได้ยาก ผมใช้โปรแกรม EViews (และโปรแกรมที่เป็นที่รู้จักอื่นๆ) อย่างไม่ต้องสงสัย แม้กระทั่งมาสอนนักศึกษาครั้งแรกรยังต้องคอยแอบกระซิบนักศึกษาให้ลองไปหาโปรแกรมในอินเทอร์เน็ตดู

พัฒนาการด้านเทคโนโลยีคอมพิวเตอร์ ตลอดจนการเติบโตของความร่วมมือในการสร้างชุมชนนักพัฒนา และใช้โปรแกรมฟรีเช่น GNU project (www.gnu.org) ได้ส่งเสริมการใช้และพัฒนาซอฟต์แวร์ฟรี ซึ่งหมายถึงอิสระภาพในการใช้ สำเนา แจกจ่าย ศึกษาตลอดจนการปรับปรุงโปรแกรมต่างๆ ภายใต้ GNU ก่อให้เกิดโปรแกรมฟรีจำนวนหนึ่งที่สามารถนำมาใช้ในการวิเคราะห์ข้อมูล

หนึ่งในโปรแกรมด้านการวิเคราะห์ข้อมูลที่เป็นที่นิยมมากที่สุด ในหมู่นักวิชาการและภาคเอกชนทั่วโลก คือ โปรแกรมอาร์ (R) (ดูรายละเอียดได้ที่ www.r-project.org) โดยโปรแกรมดังกล่าวเกิดจากการริเริ่มของศาสตราจารย์ Robert Gentleman และศาสตราจารย์ Ross Ihaka ซึ่งเป็นศาสตราจารย์ ด้านสถิติแห่ง

มหาวิทยาลัยโอคแลนด์ นิวซีแลนด์ โดยมีเป้าหมายที่จะพัฒนาโปรแกรมฟรีให้นักศึกษาได้ใช้ ทั้งสองได้รับแรงบันดาลใจจากโปรแกรมภาษา S ซึ่งถูกพัฒนาโดย John Chambers และผู้ร่วมงานที่ Bell Labs



ภาพประกอบจากอินเทอร์เน็ต

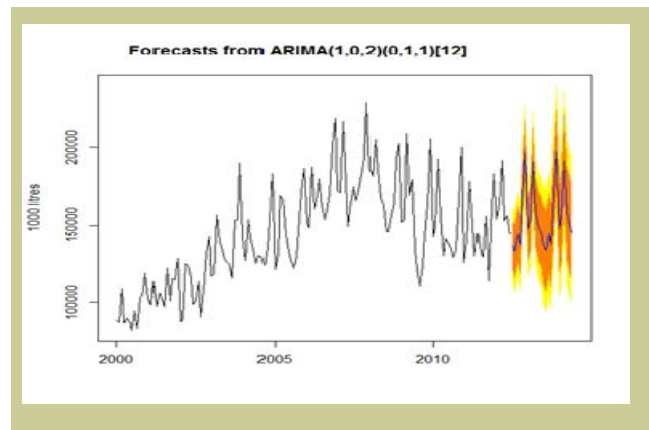
เดิมหน่วยงานทั้งหลายมักพึ่งพิงซอฟต์แวร์ที่ต้องจ่ายค่าลิขสิทธิ์เช่น SAS (www.sas.com) และ SPSS (<http://www-01.ibm.com/software/analytics/spss/>) อย่างไรก็ตาม ปัจจุบันบริษัทข้ามชาติยักษ์ใหญ่ทั้งหลายเช่น Google, Facebook, Pfizer, Bank of America, Merck, InterContinental Hotels Group, Shell, Motorola ได้หันมาใช้โปรแกรมฟรีในการวิเคราะห์ข้อมูลเพื่อใช้ในการกำหนดราคา หรือในการทดสอบทางคลินิก หรือใช้ในการสร้าง และปรับเปลี่ยนจำลองทางการเงิน ตัวอย่างเช่น Facebook ใช้โปรแกรมอาร์ในการวิเคราะห์ว่า

ข้อมูลใดจะช่วยทำนายว่าผู้สมัครใช้เฟสบุ๊คจะใช้บริการต่อไป หรือยังคงใช้อยู่เมื่อเวลาผ่านไป

โดยพื้นฐานแล้ว โปรแกรมอาร์ เป็นโปรแกรมภาษาที่ใช้ในการเขียนโปรแกรมเช่นเดียวกับโปรแกรมภาษา C, Java และ Perl ซึ่งทำให้ผู้ใช้สามารถใช้โปรแกรมอาร์ได้หลากหลาย โดยใช้คำสั่งต่างๆกันไป หรือสามารถเขียนคำสั่งขึ้นมาใช้เองได้ นักสถิติสามารถใช้ประโยชน์จากโปรแกรมอาร์ ในการจัดการข้อมูล การวิเคราะห์ข้อมูล ตลอดจนการนำเสนอข้อมูลด้วยแผนภาพต่างๆ ผู้อ่านอาจจะสงสัยว่าเราจำเป็นต้องเขียนโปรแกรมทั้งหลายเองหรือไม่ในการวิเคราะห์ข้อมูลต่างๆ คำตอบคือไม่จำเป็น เนื่องจากการเป็นโปรแกรมฟรี ส่งผลให้นักสถิติและผู้ใช้โปรแกรมอาร์จำนวนมากเขียนชุดคำสั่ง (package) ซึ่งคือกลุ่มของคำสั่งที่ใช้ตอบคำถามหรือวิเคราะห์เรื่องเดียวกัน เช่น ชุดคำสั่งในการวิเคราะห์ข้อมูลที่เป็นอนุกรมเวลา (zoo หรือ tseries) ชุดคำสั่งสำหรับการวิเคราะห์สำหรับการวิเคราะห์ทางการเงิน (RQuantLib สำหรับวิเคราะห์ราคาอนุพันธ์, portfolio สำหรับการวิเคราะห์พอร์ต)

บริษัทห้างร้านต่างๆที่มีข้อมูลยอดขายของบริษัทตัวเองก็สามารถที่จะนำข้อมูลดังกล่าวเพื่อมาคาดการณ์ยอดขาย และวางแผนการผลิต ด้วยแบบจำลองอนุกรมเวลา(แบบจำลองที่ใช้ข้อมูลพฤติกรรมตัวเองในอดีตเป็นตัวทำนายอนาคต) เพียงนำข้อมูลเข้าไปใส่ในโปรแกรมอาร์แล้วเลือกชุดคำสั่งที่สามารถใช้ในการประมาณค่าและทำนายได้ซึ่งมีอยู่หลากหลายชุดคำสั่ง เช่นชุดคำสั่งที่ผมแนะนำคือ “forecast” ซึ่งมีแบบจำลองอนุกรมเวลาหลายประเภทให้เลือก ซึ่งแบบจำลองหนึ่งที่เป็นที่นิยมคือ แบบจำลอง Autoregressive Integrated Moving Average (ARIMA) ซึ่งในชุดคำสั่งดังกล่าวได้ผนวกคำสั่งที่เลือกแบบจำลองที่เหมาะสมให้ด้วยว่าควรจะใช้ข้อมูลย้อนหลังเท่าไรควรจะมีผลของฤดูกาลเข้าไปด้วยหรือไม่เพื่อให้ผู้อ่านเห็นภาพว่า ผลที่ได้จากการวิเคราะห์จะเป็นเช่นไร ผมได้ลองใช้ชุดคำสั่งดังกล่าวทำนายปริมาณการบริโภคเบียร์ของคนไทย (ข้อมูลจาก www.bot.or.th) แล้วนำมาวาดเป็นรูปจะเห็นแนวโน้มการบริโภคเบียร์ในอนาคต ดังรูปข้างล่าง ซึ่งเห็นได้ว่าการบริโภคเบียร์มีลักษณะ

เป็นฤดูกาล โดยเส้นสีฟ้าคือค่าทำนายและมีแถบความเชื่อมั่นของการทำนายอยู่ด้วย



ที่มา: การคำนวณของผู้เขียน

เมื่อเปรียบเทียบกับโปรแกรมเพื่อการคำนวณอื่น ๆ จะเห็นได้ว่า โปรแกรมอาร์มีข้อได้เปรียบ คือ ราคาถูกกว่า เนื่องจากเป็นโปรแกรมฟรี ซึ่งสามารถลงโปรแกรมได้ในคอมพิวเตอร์ทุกเครื่องที่บริษัทมี ไม่ว่าจะทำงานให้บริษัทเล็กหรือใหญ่ ก็สามารถจะมีโปรแกรมดังกล่าวไว้ใช้ได้ หรือนักศึกษาที่กำลังเริ่มทำวิจัย ในขณะที่โปรแกรมเพื่อการคำนวณที่ต้องเสียค่าลิขสิทธิ์ในราคาหลายหมื่นบาท อีกทั้งมีความเสี่ยงที่จะต้องปรับเวอร์ชันใหม่อยู่บ่อยๆ นอกจากนี้ การลงโปรแกรมสามารถทำได้ง่ายตามคำแนะนำในเว็บไซต์ นอกจากนี้โอเพนซอร์สซอฟต์แวร์ยังมีข้อได้เปรียบที่ผู้ใช้สามารถตรวจสอบความถูกต้องของโค้ดคำสั่งต่างๆ อีกทั้งยังมีคู่มือสำหรับการใช้แต่ละชุดคำสั่ง

นอกจากเครื่องมือใหม่ๆทางเศรษฐมิติ หรือทางการเงินที่ผู้อ่านต้องการใช้ ผู้อ่านไม่ต้องรอให้บริษัทผลิตโปรแกรมสำเร็จรูปสร้างคำสั่งมา เพราะชุมชนผู้ใช้โปรแกรมอาร์มักจะสร้างคำสั่งออกมาอย่างรวดเร็ว นอกจากนี้ เนื่องจากนักสถิติหรือนักวิชาการทางการเงินจำนวนหนึ่งใช้โปรแกรมอาร์ในการทำวิจัยอยู่แล้ว ดังนั้นคำสั่งดังกล่าวก็สามารถนำมาประยุกต์กับข้อมูลที่เราได้ทันที ตัวอย่างเช่น แบบจำลองที่ชื่อ ARFIMA (Fractional Integrated ARMA) ซึ่งเริ่มถูกนำมาใช้ในการวิเคราะห์อนุกรมเวลาในปลายทศวรรษที่ 1990 ด้วยโปรแกรมภาษาที่นักวิจัยต้องเขียนขึ้นเอง แต่โปรแกรมสำเร็จรูปเช่น EViews หรือ Stata ฟังผนวกคำสั่งสำหรับแบบจำลองดังกล่าวในโปรแกรมเมื่อไม่นานมานี้ ทั้งๆ ที่ชุมชนผู้ใช้อาร์ได้ดำเนินการมานานพอสมควรแล้ว (ชุดคำสั่ง fArma)

ปัจจุบันจากเว็บไซต์ของโครงการอาร์ ได้ระบุว่า มีชุดคำสั่งอยู่ 3,973 ชุดคำสั่ง ให้เลือกใช้ได้ตามความต้องการของหน่วยงาน ในปีที่ผ่านมา ผมได้มีโอกาสสอนในวิชาเศรษฐมิติทางการเงินเบื้องต้นที่เน้นด้านอนุกรมเวลา ผมได้ใช้โปรแกรมอาร์เป็นโปรแกรมหลักซึ่งแน่นอนโปรแกรมอาร์สามารถใช้ประมาณค่าแบบจำลองทุกแบบจำลองที่นักศึกษาจำเป็นต้องทราบ และนำไปใช้ประโยชน์ต่อไป

ผู้ที่ไม่เคยใช้โปรแกรมที่มีลักษณะการทำงานบน command line อาจจะไม่ชอบการใช้โปรแกรมอาร์ อย่างไรก็ตาม นักพัฒนาซอฟต์แวร์ได้พัฒนาชุดคำสั่งเช่น Rcmdr สำหรับการทำงานในลักษณะ Graphic User Interface(GUI) ควบคู่ไปกับโปรแกรมอาร์ ดังนั้นผู้ใช้สามารถนำเข้าข้อมูล วิเคราะห์ข้อมูลด้วยสถิติพื้นฐาน แสดงกราฟ และวิเคราะห์แบบจำลองเบื้องต้นบน GUI

ผู้อ่านอาจสงสัยว่าความนิยมในโปรแกรมอาร์มีมากน้อยแค่ไหน Ashlee Vance เขียนคอลัมน์ในหนังสือพิมพ์ New York Times เมื่อวันที่ 7 มกราคม 2552 ว่าผู้ที่คุ้นเคยกับโปรแกรมประมาณการไว้ว่าน่าจะมีผู้ใช้โปรแกรมอาร์เป็นประจำประมาณ 250,000 ล้านคน ความนิยมของโปรแกรมอาร์ที่เพิ่มขึ้นเป็นภัยคุกคามต่อสถาบัน SAS ซึ่งเป็นกิจการที่เน้นการวิเคราะห์ข้อมูลซึ่งมีรายได้ต่อปีมากกว่า 2 พันล้านเหรียญสหรัฐ ซึ่งเดิมเคยเป็นโปรแกรมที่นักวิชาการตลอดจนผู้จัดการบริษัททั้งหลายใช้ (เนื่องจากคณะด้านธุรกิจส่วนมากมักจะใช้โปรแกรม SAS) แต่ปัจจุบัน Kuhn นักสถิติที่บริษัท Pfizer พบว่า นักศึกษาจบใหม่ระดับบัณฑิตศึกษาในปัจจุบันมักจะจบมาพร้อมกับอาร์เป็นภาษาที่สอง

ในแง่ของสาขาการเงิน ความนิยมการเงินเชิงคำนวณ (computational finance) ในหลายปีที่ผ่านมาส่งผลต่อความต้องการโปรแกรมในการวิเคราะห์ข้อมูลตลอดจนการสร้างแบบจำลองด้วยโปรแกรมคอมพิวเตอร์ เนื่องจากต้นทุนที่เป็นตัวเงินที่ต่ำของการใช้โปรแกรมอาร์ จึงเริ่มมีโครงการบัณฑิตศึกษาด้านการเงินเชิงคำนวณในสหรัฐอเมริกาที่มีโครงสร้างหลักสูตรที่เกือบทุกวิชาสอนควบคู่โดยใช้โปรแกรมอาร์โดยเฉพาะ เช่นที่มหาวิทยาลัยแห่งรัฐวอชิงตัน (<http://computational-finance.uw.edu>) ในขณะที่ช่วงไม่กี่ปีที่ผ่านมา มีหนังสือที่เกี่ยวข้องกับโปรแกรมอาร์ออกมาหลายสิบเล่ม

ในโลกเศรษฐกิจปัจจุบันที่เราได้รับข้อมูลจำนวนมาก ทั้งในระดับกิจการและระดับมหภาค ผมคิดว่าการตัดสินใจและการวางแผนต่างๆน่าจะมีพื้นฐานจากข้อมูลเหล่านั้น ด้วยเครื่องมือฟรีที่ผมได้แนะนำไว้ในบทความนี้น่าจะเป็นจุดเริ่มต้นที่ดีที่ผู้อ่านจะลองใช้เครื่องมือฟรีดังกล่าวเพื่อประโยชน์ในการดำเนินธุรกิจและงานวิจัยต่างๆ



คอลัมน์เศรษฐศาสตร์

มีวัตถุประสงค์เผยแพร่องค์ความรู้ด้านเศรษฐศาสตร์สู่สาธารณะ โดยได้รับการสนับสนุนจากหนังสือพิมพ์ประชาชาติธุรกิจ ตีพิมพ์บทความทุก 4 สัปดาห์ ท่านที่สนใจอ่านบทความย้อนหลังได้ที่ www.econ.tu.ac.th

ประสานงานและดำเนินการโดย

คณะทำงานสัมมนาและเผยแพร่

คณะเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์

เลขที่ 2 ถนนพระจันทร์

พระนคร กรุงเทพฯ 10200