

SỰ KẾT NỐI THỐNG KÊ KINH TẾ VỚI MỌI NGƯỜI: VAI TRÒ CHO CÁC NGUỒN DỮ LIỆU THAY THẾ*?

Zeynep Orhun Girard, Phòng Thống kê ESCAP

Tóm tắt

Để thống kê chính thức có thể triết xuất được các giá trị phân tích từ dữ liệu lớn, thì cần phải được hướng dẫn chặt chẽ bởi các nhà hoạch định chính sách thống kê với mục tiêu phương pháp luận thực tế và thu hẹp các khoảng trống dữ liệu giữa các lĩnh vực khác nhau. Để minh họa điều này, bài viết đã đưa ra lý luận nhằm nhấn mạnh các khía cạnh phân phối thu nhập hộ gia đình trong thống kê kinh tế bằng cách đề cập đến sự phát triển phương pháp luận thống kê kinh tế ví mô, các phương pháp tiếp cận đầy đủ số liệu thống kê, các sáng kiến liên quan đến chính sách như công việc của Ủy ban Stiglitz Sen Fitoussi, khoảng trống dữ liệu của G-20 và chương trình nghị sự phát triển sau năm 2015. Bài viết thảo luận về nguồn dữ liệu lớn có tiềm năng để tạo điều kiện cho các mối liên kết giữa các số liệu thống kê kinh tế với mọi người.

Xem xét các nghiên cứu hiện có sử dụng tìm kiếm dữ liệu trực tuyến thống kê với dữ liệu lớn trong những lĩnh vực khác nhau của thống kê kinh tế, bài viết đưa ra đề xuất mà dữ liệu lớn có thể cải thiện: (i) chất lượng và mức độ chi tiết các dữ liệu hiện có để có thể “Đúng thời điểm cần thiết và lấy con người làm trung tâm hoạch định chính sách”; (ii) Sự sẵn có của dữ liệu cần thiết mà không thể thu thập được thông qua các phương tiện truyền thông, được thảo luận thông qua ví dụ hàng tiêu dùng. Các kết luận thảo luận về những thách thức của việc sử dụng dữ liệu lớn cho thống kê chính thức và các ý tưởng nghiên cứu thêm.

“Gió không thuận anh ấy không thể đến được bên bờ của số phận”.

Michel de Montaigne

1. Giới thiệu

Montaigne sống lúc chưa có sự xuất hiện của các nguồn dữ liệu thay thế trên thế giới nhưng những cảnh báo của ông rằng kê chưa được quan tâm, nhưng khi đặt trong bối cảnh của chương trình nghị sự phát triển sau năm 2015 với những tài liệu tham khảo phong phú về cuộc cách mạng dữ liệu và các nguồn dữ liệu thay thế. Đặc biệt cảnh báo dữ liệu lớn giúp chúng ta biết rằng chúng ta biết cần

phải làm gì nếu muốn tránh bị mất dữ liệu trong trận ngập lụt dữ liệu.

Bài viết tập trung vào cơ hội và thách thức của dữ liệu lớn đối với tương lai của thống kê kinh tế. Quan điểm nghiên cứu tiếp tục ở đây là dữ liệu lớn và thống kê kinh tế có ý nghĩa như thế nào nếu nó được định hướng chặt chẽ bởi chính sách thống kê, với phương pháp luận thực tế và thu hẹp những khoảng trống dữ liệu. Trong thống kê kinh tế, những phát

* Lưu ý chung: Các quan điểm trình bày của tác giả không nhất thiết phản ánh quan điểm và vị trí của Liên hợp quốc

triển mới cả về phương pháp thống kê kinh tế vĩ mô và đo lường cần chỉ rõ đến một nhu cầu liên kết số liệu thống kê kinh tế với những người bị ảnh hưởng. Nguồn dữ liệu thay thế có thể có một vai trò tiềm năng để tạo liên kết bằng cách tăng cường các khía cạnh phân phối số liệu thống kê kinh tế và dữ liệu sẵn có. Nói cách khác, dữ liệu lớn có thể bổ sung thêm cho nguồn dữ liệu "truyền thống" bằng cách tăng mức độ chi tiết và bảo đảm (cả về không gian và thời gian) chứ không phải là thay thế dữ liệu "truyền thống".

Bài viết đưa ra những lý luận về tương lai của thống kê kinh tế và nguồn dữ liệu lớn nhằm nhấn mạnh những khía cạnh phân phối và thu nhập hộ gia đình trong thống kê kinh tế vĩ mô. Phần 2 tóm tắt "Quá trình thay đổi hoàn thiện" các khung thống kê kinh tế vĩ mô và các cuộc thảo luận toàn cầu liên quan đến tác động của số liệu thống kê kinh tế đối với phúc lợi của người dân và hoạch định chính sách. Phần này tiếp tục đề cập đến việc so sánh nguồn dữ liệu kinh tế vĩ mô và vĩ mô được OECD thực hiện như là một công việc theo dõi của Ủy ban Stiglitz Sen Fitoussi. Phần 3 xem xét một số ví dụ về cách thức sử dụng dữ liệu lớn để giải thích các hiện tượng kinh tế, đặc biệt là trong các hình thức sử dụng truy vấn dữ liệu trực tuyến. Phần 4 đề xuất cách thức có thể sử dụng dữ liệu lớn để đưa số liệu thống kê kinh tế đến mọi người bằng cách lấy hàng tiêu dùng làm ví dụ. Phần 5 kết luận với một số thách thức và ý tưởng để nghiên cứu thêm.

2. Chúng ta muốn gì cho thống kê kinh tế trong tương lai?

Khi một người không chuyên về thống kê được yêu cầu cung cấp ví dụ về chỉ tiêu kinh tế, nhiều khả năng câu trả lời sẽ là tổng sản phẩm trong nước (GDP). GDP là hình ảnh thu nhỏ của các

số liệu thống kê kinh tế cho nhiều người; là bản chụp về cách đo sức mạnh nền kinh tế của một quốc gia, là chỉ số chính của Hệ thống tài khoản quốc gia (SNA) và là khung tích hợp cho tất cả các số liệu thống kê kinh tế và nhiều hơn thế³.

Sự ảnh hưởng của tài khoản quốc gia và lý thuyết thống kê kinh tế cơ bản đối với các dữ liệu chúng ta sản xuất và số liệu thống kê kinh tế chúng ta biên soạn không thể được phóng đại. Tương tự như vậy, những thay đổi trong nền kinh tế chính trị toàn cầu trong suốt vài thập kỷ qua đã trực tiếp ảnh hưởng đến các loại và số lượng dữ liệu về nền kinh tế của hệ thống thống kê quốc gia đã sản xuất.

Sự phát triển và sửa đổi các khung thống kê kinh tế vĩ mô cũng như lý thuyết thống kê kinh tế⁴ vẫn đang tiếp tục và chắc chắn sẽ còn tiếp tục theo thời gian. Loạt mới nhất các khung sửa đổi liên quan đến bảng cân cán thanh toán và đầu tư quốc tế (BPM6), Hệ thống tài khoản quốc gia (SNA 2008) và Thông kê Tài chính Chính phủ (GFS 2014), đang tiếp tục được nỗ lực hoàn thiện và hài

³ Điều này không có nghĩa là mọi người nhận thức được những gì thực sự quan trọng đòi hỏi các chỉ số kinh tế. Dựa trên các cuộc khảo sát tiến hành của OECD tại 27 quốc gia EU, Thổ Nhĩ Kỳ và Croatia năm 2007, chỉ ra rằng hơn một nửa số người dân châu Âu không có một ý tưởng gì về tỷ lệ tăng trưởng GDP và chỉ có 8% biết chính xác con số này. Xem thêm Giovannini, Enrico. "Đưa Thống kê đến Người dân: A must to build democracy in the XXI century". Có sẵn tại: <http://www.unece.org/stats/documents/ece/ces/bur/2008/25e.pdf>

⁴ Ở châu Á và Thái Bình Dương, các thiết lập cơ bản của thống kê kinh tế mà một hệ thống thống kê quốc gia cần phải có khả năng sản xuất được thể hiện trong bảy thành phần thiết lập cốt lõi: giá cả và chi phí, nhu cầu và sản lượng, thu nhập và giàu có, tiền bạc và ngân hàng, chính phủ, thị trường lao động, tài nguyên thiên nhiên và môi trường. Vui lòng xem E/ESCAP/CST(2)/4 "Proposed Core Set of Economic Statistics for Asia and the Pacific. Note by the ESCAP Technical Advisory Group on the Development Economic Statistics". ESCAP Committee on Statistics. 15-17 December 2010

hòa ở cấp độ khái niệm để phản ánh chính xác hơn nền kinh tế chúng ta đang sống. SNA 2008, BPM6 và GFS 2014 áp dụng các khái niệm tương tự cho cư trú (dựa trên lợi ích các tổ chức kinh tế chủ yếu), ghi vào hóa đơn xuất khẩu và nhập khẩu (dựa trên thay đổi của thành phần kinh tế), tích lũy tài sản so với tài sản trên cơ sở tiền mặt và bảng cân cân thanh toán⁵. Các quốc gia trên toàn thế giới hiện đang trong quá trình triển khai thực hiện hoặc có kế hoạch thực hiện các phiên bản mới nhất của các khung thông kê kinh tế vĩ mô.

Sự phát triển của Khung thông kê kinh tế vĩ mô được mô tả ở trên chỉ ra sự thay đổi hướng tới sự lựa chọn bao gồm các hoạt động phi thị trường và tập trung vào các khía cạnh kinh tế xã hội tốt của hoạt động kinh tế, ví dụ sự phân phối giữa các tổ chức khác nhau cơ bản phù hợp, vì có tính đến những tác động của toàn cầu hóa nền kinh tế quốc gia. Một hướng tương tự cũng được nhắc lại trong 20 kiến nghị đưa ra sáng kiến khoảng trống dữ liệu của G-20, những khuyến nghị 15-20 không thuộc ngành tài chính và kinh tế mà bộ dữ liệu được bảo đảm, ví dụ, phát triển của một chiến lược để biên dịch và phổ biến bảng cân đối, dòng vốn, dữ liệu ngành và việc biên soạn các thông tin về phân phối thu nhập, ví dụ: phạm vi và tứ phân vị (Eurostat)⁶. Bảng cân đối kinh tế hộ gia đình, trong đó chứa dữ liệu đo lường "tổng giá trị tài sản các hộ gia đình và nguồn

vốn, bao gồm cả giá trị thực, là phần không thể thiếu của một hệ thống các tài khoản hoàn chỉnh" (OECD 2013). Tuy nhiên, một khoảng trống dữ liệu quan trọng cho việc lập bảng cân đối liên ngành quan tâm, cụ thể là khu vực phi tài chính, đặc biệt là khu vực doanh nghiệp phi tài chính và hộ gia đình. Đây là thông tin chi tiết cần thiết cho việc liên kết các thành phần thực và tài chính, cũng như cho phép phân tích rủi ro và mức độ bị tổn thương của nền kinh tế (Leone 2009).

Tăng cường tập trung cả vào các khía cạnh phân phối thu nhập và xây dựng bảng cân đối cần thêm dữ liệu phân tích mà chúng ta không thể hiện được, như phạm vi khu vực của cuộc điều tra mẫu thường không đạt được mức độ mong muốn của phân tích và các dữ liệu thay thế, ví dụ: dữ liệu thu được từ các cuộc tổng điều tra hoặc hồ sơ hành chính, có kích thước mẫu thích hợp cho phân tích chi tiết nhưng có thể không đo lường được các hiện tượng một cách thỏa đáng.

Trong khi các phương pháp luận về "Quá trình thay đổi hoàn thiện" đã được tính tới, thì một hướng phát triển khác liên quan đến số liệu thông kê kinh tế vĩ mô đã tập trung vào việc tìm hiểu tác động của nó đối với việc ra quyết định. Mặc dù nguồn gốc của tất cả các số liệu thông kê luôn luôn là do cá nhân trả lời cho chính bản thân họ hoặc trên danh nghĩa là một thành viên hộ gia đình trong điều tra lực lượng lao động hoặc cho doanh nghiệp họ làm việc trong điều tra doanh nghiệp, như vậy các khung thông kê kinh tế vĩ mô sản xuất số liệu thông kê kinh tế tổng hợp che giấu sự khác biệt quan trọng giữa các nhóm người. Ví dụ cho thấy sự không liên kết giữa số liệu kinh tế vĩ mô tổng hợp và số liệu phúc lợi kinh tế của người dân, ví dụ, so sánh xu hướng của GDP đối với thu nhập ròng quốc

⁵ Phương pháp bảng cân đối không phải là một ý tưởng mới trong thực tế. Xin vui lòng xem hướng dẫn bản quốc tế đầu tiên về chủ đề này "Provisional international guidelines on the national and sectoral balance-sheet and reconciliation accounts of the systems of national accounts". Statistical Papers. Series M, No. 60. United Nations, New York, 1977

⁶ Vui lòng xem Eurostat. "G20 Data Gaps Initiative (DGI) -background". Có sẵn tại:

[http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/G20_Data_Gaps_Initiative_\(DGI\)_%E2%80%93background](http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/G20_Data_Gaps_Initiative_(DGI)_%E2%80%93background)

gia và thu nhập thực tế của hộ gia đình được coi là phù hợp cho việc chỉ ra phúc lợi kinh tế của các hộ gia đình.

Tất cả báo cáo của Ủy ban Stiglitz Sen Fitoussi năm 2008 đã được công bố công khai các cuộc thảo luận chủ yếu đặt câu hỏi về các phép đo lường chúng ta sử dụng để đánh giá các phúc lợi kinh tế của người dân và trong đó nhấn mạnh các giới hạn của số liệu thống kê kinh tế vĩ mô. Báo cáo đưa ra năm khuyến nghị cải cách thống kê kinh tế để phản ánh tốt hơn đời sống của người dân (Stiglitz et al 2008):

Khuyến nghị 1: Thu nhập và tiêu dùng thực là những chỉ số chính xác hơn về phúc lợi kinh tế so với sản xuất (thể hiện trong GDP), nó tùy thuộc vào sự thay đổi dựa trên giá trị hao mòn, thu nhập ròng, và sự khác biệt giữa giá cơ bản (mà tại đó đầu ra có giá trị) và giá thị trường (mà tại đó hàng tiêu dùng có giá trị).

Khuyến nghị 2: Nhấn mạnh quan điểm hộ gia đình bao gồm cả chuyển đổi loại hình xã hội mong muốn.

Khuyến nghị 3: Có một nhu cầu cho việc phân tích kết hợp giữa thu nhập, tiêu dùng và sự tích lũy tài sản của hộ gia đình, ví dụ như thông qua bảng cân đối kinh tế hộ gia đình.

Khuyến nghị 4: Con số trung bình ẩn chứa xu hướng mức sống đa dạng hơn là nhiều dữ liệu cần thiết để có được phân phối thu nhập, tiêu dùng và sự giàu có.

Khuyến nghị 5: Hàng hóa và dịch vụ của sản xuất hộ gia đình cho riêng tiêu dùng cuối cùng phải được đưa vào tài khoản.

Kể từ đó, Tổ chức Hợp tác kinh tế và phát triển (OECD) đã phát triển các phương pháp cho công

việc của mình rộng hơn và tiên hành các phương pháp đó với các quốc gia hàng đầu như Úc⁷. Như vậy đã làm sáng tỏ về tương lai chúng ta muốn cho thống kê kinh tế, đó là việc đưa ra các quyết định ảnh hưởng nhưng cũng có thể che giấu đi rất nhiều sự ảnh hưởng đó đến cuộc sống của họ. Lưu ý rằng sức ép giữa các chỉ tiêu tài khoản quốc gia và mong muốn để sử dụng chúng của các nhà làm chính sách như là các chỉ số phúc lợi đã được sử dụng liên tục trong nhiều thập kỷ. Minh họa đầu tiên là nghiên cứu đo lường phúc lợi kinh tế của Nordhaus và Tobin để mở rộng thu nhập quốc gia vào một chỉ số phúc lợi (Nordhaus và Tobin 1972).

Các phiên bản của khung thống kê kinh tế vĩ mô và điều tra thêm về các giới hạn của số liệu thống kê kinh tế vĩ mô có được từ khi tập trung trong một cách tiếp cận đầy đủ số liệu thống kê kinh tế, xã hội và môi trường mà còn được dùng trong các chương trình nghị sự phát triển bền vững mới nổi gần đây. Báo cáo Hội đồng cấp cao trong chương trình nghị sự phát triển sau năm 2015 đề cập đến cuộc cách mạng dữ liệu cho thấy sự cần thiết phải nâng cao chất lượng số liệu thống kê⁸, sử dụng công nghệ mới, phân tích chi tiết hơn nữa (theo giới tính, địa lý, thu nhập, v.v..) cho hoạch định chính sách và mục tiêu cụ thể tác động tới xã hội, môi trường kinh doanh và tăng trưởng kinh tế.

⁷ Đặc biệt, OECD lập một nhóm chuyên gia đến làm việc về sự bất bình đẳng trong tài khoản quốc gia và sản xuất đồng thời hướng dẫn sử dụng hai Khung thống kê của OECD về việc phân phối thu nhập của hộ gia đình, tiêu thụ và sự tích lũy tài sản và hướng dẫn của OECD cho Thống kê vi mô về hộ gia đình giàu có.

⁸ Có rất nhiều các khung đảm bảo chất lượng được thiết kế bởi các tổ chức quốc tế được sử dụng trong thống kê chính thức; thường với các thành phần chồng chéo như tính liên quan, chính xác, kịp thời, đúng giờ, tiếp cận, rõ ràng, và so sánh và tính nhất quán theo thời gian. Chi phí hiệu quả cũng có thể được thêm vào như là một tiêu chí về chất lượng.

Ai đó có thể giải thích điều này được gọi là một cầu nối tất cả các lĩnh vực chính của thống kê, bao gồm liên kết số liệu thống kê kinh tế với những người mà họ bị ảnh hưởng.

Do đó, một mặt, các nhà thống kê kinh tế đã làm việc để hướng tới một hệ thống thống kê kinh tế vĩ mô nội bộ chặt chẽ hơn, do sự thay đổi cơ cấu mà còn bởi chu kỳ khủng hoảng kinh tế và tài chính, kết quả là các dự án như các sáng kiến khoảng trống dữ liệu G20. Mặt khác, yêu cầu chính sách mở rộng phạm vi không chỉ được xác định ít nhất bằng việc phát triển chương trình nghị sự quốc tế mới, điều đó đã chỉ ra một hướng mở cho thống kê kinh tế vĩ mô để thu hút hơn với lĩnh vực thống kê khác chủ yếu là thống kê xã hội và môi trường. Cộng đồng thống kê quốc tế đã thực hiện thử thách này bằng nhiều cách khác nhau, ví dụ đưa ra các hướng dẫn về Hệ thống môi trường kế toán (SEEA) và thêm vào các nghiên cứu về nguồn dữ liệu vi mô và vĩ mô của dữ liệu trong trường hợp phúc lợi kinh tế hộ gia đình mới nhất của một trong số đó đã được thực hiện bởi OECD. Nội dung "Xây dựng phát triển các sáng kiến hiện có để đo lường sự tiên bộ về phát triển bền vững mà bổ sung cho GDP..." vào năm 2030, trong khung các Mục tiêu phát triển bền vững là một mục tiêu (Nhóm Công tác mở năm 2014).

Tính hữu dụng và chất lượng của số liệu thống kê kinh tế liên kết ở cấp vi mô và vĩ mô

Nhóm chuyên gia OECD-Eurostat về sự bất bình đẳng trong tài khoản quốc gia đã kết hợp so sánh các nguồn dữ liệu vi mô và vĩ mô về thu nhập hộ gia đình, chi tiêu và sự tích lũy tài sản dựa trên kinh nghiệm của 21 quốc gia. Các kết quả khẳng định ước tính của tài khoản quốc gia nói chung là cao hơn so với các dự báo dựa trên dữ liệu vi mô.

Đối với thu nhập từ hộ gia đình, thu nhập từ nghề tự do liên quan; lợi tức, cổ tức nhận được; Giá chuyển đổi được quy gán bằng hiện vật và giá chuyển đổi xã hội (StiK) là những ví dụ mà khoảng trống dữ liệu được ước tính được quan sát giữa dữ liệu vi mô và vĩ mô. Ngoài lý do liên quan đến phân loại và định giá trong các ước tính vi mô và vĩ mô khác nhau, cũng có sự chênh lệch trong phạm vi bảo đảm; Ví dụ quy gán giá chuyển đổi bằng hiện vật và StiK không được quan sát thấy trong các nguồn dữ liệu vi mô. Trong tiêu dùng các hộ gia đình, nghiên cứu cho thấy chi phí về đồ uống có cồn, thuốc lá, các hàng hóa và dịch vụ (chăm sóc cá nhân, hành lý cá nhân, bảo trợ xã hội, các dịch vụ bảo hiểm và dịch vụ tài chính) là những mặt hàng mà có sự chênh lệch đáng kể giữa vi mô và vĩ mô. Khoảng trống đó được xác định thông qua việc so sánh các nguồn dữ liệu vi mô và vĩ mô cho tiêu dùng hộ gia đình liên quan đến việc điều chỉnh lãnh thổ (để điều chỉnh cho các khoản chi tiêu của các hộ gia đình không cư trú tại một quốc gia và của các hộ gia đình cư trú ở nước ngoài) mà thường chỉ tiêu được thực hiện tầm vĩ mô. Ngoại tệ và tiền gửi; cổ phiếu và vốn chủ sở hữu khác; thế chấp là những thành phần hộ gia đình giàu có, nơi khoảng trống dữ liệu đã được quan sát thấy giữa các nguồn vi mô và vĩ mô (Fesseau at al. 2013).

Nhóm chuyên gia OECD-Eurostat cũng thử nghiệm với các nguồn dữ liệu vi mô và vĩ mô để giới thiệu các khía cạnh phân phối thu nhập trong toàn bộ hệ thống tài khoản quốc gia theo khuyến nghị SNA đối với sản xuất cố định có từ thu nhập hộ gia đình, đặc điểm người tham gia trong hộ gia đình, kích thước và vị trí của các hộ gia đình. Tổng hợp các khuyến nghị được sử dụng trong nghiên cứu này là thu nhập của hộ gia đình (chính và dùng một lần), tiêu dùng cuối cùng và tiết kiệm. Nghiên cứu

này nhấn mạnh một trong những hạn chế của các nguồn dữ liệu vi mô là nguồn dữ liệu vi mô loại trừ một bộ phận dân cư, được tính đến bởi các tài khoản quốc gia (tức là những người không có địa chỉ thường trú, hộ thể chế, cư dân vùng lãnh thổ ở nước ngoài và các khu vực dân cư thưa thớt, người thu nhập cao, v v...). Thiếu sót khác của nguồn dữ liệu vi mô liên quan đến việc thiếu các thành phần, phân loại và sự chênh lệch giữa các nguồn dữ liệu vi mô và vĩ mô đã được quan sát do việc định giá giao dịch khác nhau (Fesseau et Mattonetti 2013).

Các thành phần của bảng cân đối tài sản hộ gia đình, gồm tài sản phi tài chính, tài sản tài chính và nợ⁹. Như nhận định của nhóm chuyên gia OECD về dữ liệu Thống kê vi mô thu nhập hộ gia đình, tiêu dùng và sự tích lũy tài sản, hầu hết các nước không có "đầy đủ và kịp thời số liệu bảng cân đối và điều này là đặc biệt đối với các dữ liệu về tài sản phi tài chính (OECD ICW 2013). Trong khung này, hàng tiêu dùng đang được coi là tài sản phi tài chính trong bảng cân đối, các dịch vụ từ hàng tiêu dùng được coi là tiêu thụ trong hộ gia đình, và các dịch vụ không bao gồm khấu hao và số chi phí bảo trì trong thu nhập (như là một phần thu nhập từ sản xuất hộ gia đình sử dụng dịch vụ riêng).

Hàng tiêu dùng là một thành phần quan trọng của tài sản phi tài chính, trong SNA được coi như là một hình thức chi tiêu hơn là sự tích lũy tài sản (lưu ý: dịch vụ từ hàng tiêu dùng nằm ngoài ranh giới sản xuất SNA)¹⁰. Việc xử lý các hàng tiêu dùng trong SNA có thể phần nào lý giải việc thiếu dữ liệu. Tuy

nhiên, trong phiên bản SNA 2008, nó được công nhận là "... có thể có lợi ích không nhỏ về việc có một mục ghi lại trong bảng cân đối để hiển thị các giá trị của hàng tiêu dùng" (SNA 2008, p.469). Điều này là chắc chắn, đã tăng cường các đối số cho dữ liệu hàng tiêu dùng tốt hơn, thậm chí có thể là một phần của một tài khoản vệ tinh.

Các cuộc thảo luận ở trên cho thấy hiện tại các nguồn dữ liệu kinh tế, cho dù vĩ mô hay vi mô, có thiếu sót trong việc đưa ra các thông tin về phân phối thu nhập mong muốn để cho phép phân tích chi tiết hơn, ví dụ như bảo hiểm dân số, v.v... Tương tự như vậy, có một số mặt hàng không đạt được đến một mức độ thỏa đáng trong các nguồn dữ liệu, ví dụ , hàng tiêu dùng .

Trong bối cảnh này, các câu hỏi liên quan đến dữ liệu lớn cho các nhà thống kê kinh tế có thể được yêu cầu trong hai phần:

(i) Làm thế nào dữ liệu lớn có thể tăng cường chất lượng và mức độ chi tiết của các dữ liệu hiện có để cho phép "Sản xuất đúng sản phẩm - đúng số lượng - đúng nơi - đúng thời điểm và lấy con người làm trung tâm hoạch định chính sách"?

(ii) Làm thế nào dữ liệu lớn có thể tạo điều kiện cho sự sẵn có của dữ liệu thay thế cần thiết mà không thể thu thập thông qua các phương tiện truyền thông?

"Dữ liệu lớn nên bổ sung chứ không thay thế dữ liệu nhỏ".

Rob Kitchin¹¹

⁹ "OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth". OECD, 2013

¹⁰ OECD ICW Framework treats consumer durables as non-financial assets, the services provided by those assets as household consumption, and the services net of depreciation and maintenance costs as income.

¹¹ Rob Kitchin là RC điều tra viên cao cấp về dự án lập trình Thành phố tại Viện Quốc gia về phân tích khu vực và không gian tại Đại học Maynooth, của Ireland. Vui lòng xem tại: <http://www.maynoothuniversity.ie/progcity/contributors/rob-kitchin/>

3. Dữ liệu lớn làm gì cho thống kê kinh tế đến nay?

Sử dụng dữ liệu lớn cho thống kê chính thức có thể được xây dựng như một sự hòa giải của dữ liệu đã thu thập thông qua làm chính sách cố định, hay là có sẵn như là một sản phẩm phụ của các hoạt động khác trong hệ sinh thái kỹ thuật số. Nói cách khác, chúng ta đang phải đối mặt với các dữ liệu được thu thập bởi các hệ thống thống kê quốc gia cho mục đích thống kê chính thức, các tổ chức công cộng và tư nhân như là một phần của các hoạt động của họ (ví dụ như hồ sơ y tế được tạo ra bởi các bệnh viện; tiếp thị và nghiên cứu đầu tư để tăng năng lực cạnh tranh) và với dữ liệu vừa mới ra nơi đó; bởi vì các cá nhân cố ý hoặc ngẫu nhiên sẽ thông tin (ví dụ như thông qua các phương tiện truyền thông xã hội). Điểm sáng kiến xung toàn cầu hóa của Liên hợp quốc là dựa theo các danh mục dữ liệu lớn: "xả dữ liệu; thông tin trực tuyến; cảm biến vật lý; Hồ sơ công dân hoặc dữ liệu có nguồn gốc từ một đồng"(UN Global Pulse 2012).

Do quy mô, sự phức tạp, phạm vi bảo đảm của cấu trúc cũng như các yêu tố phi cấu trúc, tốc độ của nó trong việc tạo ra / xác định và phân tích dữ liệu lớn đã được mô tả bởi 3v: volume (khối lượng), variety (đa dạng) và velocity (vận tốc) tất cả các cấp ngày càng tăng. Tiêu chí bổ sung đã được đề xuất cho việc mô tả dữ liệu lớn: đầy đủ nhất trong phạm vi, độ chi tiết, trạng thái trong việc xác định, quan hệ, linh hoạt trong các lĩnh vực và khả năng mở rộng về quy mô (The Philosophy of Data Science 2014).

Các ứng dụng tiềm năng cho dữ liệu lớn kết hợp với số liệu thống kê chính thức có thể được trình bày như sau:

(i) Tiên hành nghiên cứu phát triển thống kê tiên tiến;

(ii) Cung cấp những hiểu biết chất lượng thông qua đổi chiều dữ liệu;

(iii) Tăng cường khả năng của dữ liệu bằng cách thu hẹp khoảng cách số liệu.

Thống kê chính thức thường là một lĩnh vực mà nó nghiên cứu các hiện tượng đã xảy ra trong quá khứ và được quan sát. Đối với các nhà hoạch định chính sách thì các giới hạn của số liệu thống kê chính thức cũng được biết đến, và từ lâu họ đã sử dụng số liệu chính thức làm cơ sở để xây dựng các mô hình dự báo và dự báo để có được gợi ý trong tương lai gần, ví dụ:

Công tác dự báo GDP¹², sự thay đổi do việc chuyển đổi dữ liệu lớn mang lại bây giờ là làm thử và dự báo trở nên thực tế hơn và xử lý dữ liệu nhanh hơn, đôi khi mối tương quan nhân quả không rõ ràng. Khi dữ liệu lớn được sử dụng trong cách này, mà không có những giả thuyết và mô hình rõ ràng, mâu thuẫn với các nguyên tắc cơ bản của thống kê chính thức¹³ phát sinh lỗi nghiêm trọng cả về phương pháp.

Đánh giá các tài liệu nghiên cứu hiện có sử dụng dữ liệu lớn, đặc biệt là các dữ liệu truy vấn trực tuyến cho thấy ở một số lĩnh vực thống kê kinh tế: giá cả, thị trường nhà ở và lao động / việc làm. Bảng 1 đưa ra một bản sao của các nghiên cứu

¹² IMF công bố dự báo con số GDP trong hai năm tiếp theo ở World Economic Outlook nhưng những con số được sản xuất bởi các văn phòng quốc gia và không nhất thiết phải so sánh quốc tế. Có sẵn tại:

<http://www.imf.org/external/pubs/ft/weo/2014/01/weodat/index.aspx>

¹³ Lời mở đầu sửa đổi các nguyên tắc cơ bản của thống kê chính thức có thể được truy cập tại địa chỉ: <http://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf>

được lựa chọn, các nguồn dữ liệu, phương pháp cơ bản và kết quả. Điểm chung cho tất cả các nghiên cứu này là họ sử dụng số liệu thống kê chính thức là nguồn dữ liệu tham khảo và so sánh các xu hướng của khối lượng "Mục đích thống kê" theo từ khóa về các xu hướng của dữ liệu chính thức trong một khoảng thời gian tham chiếu. Trong thực tế, tất cả các nghiên cứu đã thử nghiệm nhưng chưa đáp ứng được với một khoảng trống dữ liệu trong thông kê chính thức.

Trong mọi trường hợp, có sự tương quan không nhỏ giữa xu hướng về thông kê và xu hướng về dữ liệu tìm kiếm trực tuyến. Ngoài ra, dựa vào các mô hình kinh tế lượng, dữ liệu tìm kiếm trực tuyến có thể dự báo những xu hướng của số liệu thống kê chính thức liên quan đến "thời điểm vượt qua khó khăn thích hợp", ví dụ: trong trường hợp các chỉ số lao động / việc làm vượt qua trong thời gian dài thì được dự báo với mức độ thấp hơn trong khi đó đối với các hoạt động tìm kiếm được thực hiện bởi người bán nhà tiềm năng trên thị trường nhà đất thì ảnh hưởng của các hoạt động tìm kiếm được quan sát sau sáu tháng. Trong cả hai trường hợp và qua các nghiên cứu chứng minh rằng dữ liệu tìm kiếm trực tuyến dự báo xu hướng trong kinh tế thì cũng hoạt động tương tự, nói cách khác, theo quan sát của thông kê chính thức trong bối cảnh này dữ liệu lớn đã được sử dụng cho các doanh nghiệp thử nghiệm.

Các dữ liệu cơ bản cho hầu hết các nghiên cứu trong Bảng 1 xuất phát đi từ sự yêu thích của công cụ Google Trends. Google Trends (GT) là các nguồn dữ liệu truy vấn trực tuyến được các nhà nghiên cứu sử dụng rộng rãi nhất. Dữ liệu GT cho phép một tìm kiếm trực tuyến cụ thể đưa ra như là

những chia sẻ của tổng số tìm kiếm trực tuyến trong thời gian đó. Nếu có đủ khối lượng tìm kiếm, xu hướng của GT có lợi thế cung cấp dữ liệu cục bộ. GT có một số quy trình điều khiển có thể giải quyết việc quá tải khối lượng tìm kiếm, ví dụ: loại bỏ các tìm kiếm bản sao của cùng một người dùng trong khoảng thời gian ngắn. GT không xác định nội dung tìm kiếm nhưng nhận biết được những tìm kiếm được lặp đi lặp lại.

Cho đến nay, hầu hết các nghiên cứu sử dụng dữ liệu GT đều tập trung vào việc so sánh dữ liệu này với dữ liệu hồ sơ hành chính hoặc số liệu thống kê chính thức để phát hiện mối tương quan chứ không phải là quan hệ nhân quả. Nhận thấy rằng các dữ liệu là một phần của các truy vấn trực tuyến liên quan đến một chủ đề (và các truy vấn không có sẵn trừ khi người dùng nhập đủ số lượng từ khóa tìm kiếm về chủ đề đó vào một thanh tìm kiếm trong Google), GT không sử dụng để tìm kiếm dữ liệu theo cấp độ. Tuy nhiên, các dữ liệu chia sẻ tại các cấp cục bộ có thể được sử dụng để có được một nhìn nhận về sự phân bố tỷ lệ các hiện tượng kinh tế trên cả nước.

Do những hạn chế của dữ liệu tìm kiếm, nên các nghiên cứu thường nhìn vào xu hướng tổng hợp. Ettredge et al. (2005), đầu tiên chỉ tình cờ nghiên cứu, sử dụng dữ liệu thất nghiệp chính thức chia theo giới tính và độ tuổi, phát hiện ra truy vấn dữ liệu trực tuyến được kết hợp với tình trạng thất nghiệp đối với nam giới tuổi từ 20 trở lên, trong khi đó không ai sử dụng các phép hồi quy số lượng người thất nghiệp với nữ giới cho kết quả đáng kể. Nhiều nghiên cứu cần được thực hiện để khám phá sức mạnh phân tích chi tiết các dữ liệu tìm kiếm trực tuyến và dữ liệu lớn khác cho thông kê kinh tế.

Bảng 1. Ví dụ về tìm kiếm dữ liệu trực tuyến cho thông kê kinh tế

Phạm vi Thông kê kinh tế	Nghiên cứu	Nguồn dữ liệu	Phương pháp/ kết quả
Giá tiêu dùng (thực phẩm, đồ uống và sản phẩm gia dụng)	MIT: Dự án Tỷ giá (liên tục)	Lọc Dữ liệu từ các nguồn trực tuyến được công bố công khai bao gồm các siêu thị lớn nhất trong quốc gia / chỉ số CPI chính thức	Lập chỉ số CPI và lạm phát /thực hiện trực tuyến hàng tháng và hàng năm để theo dõi chặt chẽ các chỉ số chính thức tại các nước thí điểm; có một trường hợp Argentina có khả năng là ở vấn đề chất lượng.
Thị trường nhà đất (mục đích thống kê dữ liệu là: các đại lý bất động sản)	Ngân hàng Anh: McLaren và Schanbhogue (2011)	Google Insights và Home Builders Federation, Royal Institution of Chartered Surveyors	Mô hình tự hồi quy / tìm kiếm nhu cầu cao hơn nên có một tương quan tốt giữa các tìm kiếm trực tuyến và giá nhà đất.
Thị trường nhà đất (mục đích thống kê dữ liệu: dữ liệu thực từ đơn vị bất động sản và các bất động sản)	Wu và Brynjolfsson (2009)	Google Trend và hiệp hội thương mại NAR (National Association of Realtors); Văn phòng Liên bang Giám sát giá nhà đất (nguồn số liệu thống kê chính thức của US Census Bureau)	Mô hình tự hồi quy mùa vụ / hồi quy tuyến tính doanh số bán hàng dựa trên khối lượng từ tìm kiếm & Sự tương quan giữa chỉ số giá nhà ở (HPI) và chỉ số tìm kiếm / tìm kiếm tương quan thông qua đại lý bất động sản với doanh số bán nhà ở & HPI có thể đi lên hoặc đi xuống tùy thuộc vào việc người bán hoặc người mua tìm người bán thay đổi đầu tiên và cũng phát hiện mối tương quan giữa doanh số bán nhà và hoạt động tìm kiếm cho hàng tiêu dùng trong một khoảng thời gian
Lao động / việc làm (mục đích thống kê dữ liệu: JSA - Phụ cấp công việc của Seeker)	Ngân hàng của nước Anh: McLaren và Schanbhogue (2011)	Google Insights và số lượng người khiêu nại, câu hỏi về niềm tin tiêu dùng thay đổi trong tỷ lệ thất nghiệp dự kiến của GfK; LFS	Mô hình tự hồi quy / Thay đổi việc làm trong quá khứ 3 tháng / JSA tìm kiếm cải thiện phù hợp và mang lại sức mạnh giải thích trong việc giải thích những thay đổi trong công việc và trong giờ làm thứ. JSA được tốt hơn và nhanh hơn so với GfK.
Lao động / việc làm (mục đích thống kê dữ liệu: Việc làm)	D'Amuri và Marcucci (2009)	Chỉ mục tìm kiếm của Google và bộ dữ liệu ban đầu từ Sở Lao động	Mô hình tuyến tính ARMA / chỉ mục tìm kiếm của Google chiếm ưu thế hàng đầu như là chỉ số để dự báo tỷ lệ thất nghiệp của Mỹ.

Phạm vi Thông kê kinh tế	Nghiên cứu	Nguồn dữ liệu	Phương pháp/ kết quả
Lao động / việc làm (mục đích thông kê dữ liệu: văn phòng hoặc cơ quan trợ cấp thất nghiệp; tỷ lệ thất nghiệp, nhân viên tư vấn và công việc phổ biến nhất trong công cụ tìm kiếm tại Đức)	Askitas (2009)	Google Insights và dữ liệu Cơ quan Lao động Liên bang	Chuỗi thời gian mô hình sửa sai quan hệ nhân quả / Kết quả dựa trên các dữ liệu của tuần thứ ba và thứ tư của tháng trước đó chứng minh được ý nghĩa thống kê, ví dụ: tìm kiếm cho cơ quan việc làm có tác động tích cực đến tình trạng thất nghiệp và hoạt động tìm kiếm công việc sẽ dự báo sự suy giảm mạnh mẽ và đáng kể trong tỷ lệ thất nghiệp ngắn hạn.
Lao động / việc làm (mục đích thông kê dữ liệu: tìm kiếm công việc, việc làm, www.monster.com, sơ yếu lý lịch, danh sách nơi làm việc và công việc)	Ettredge et al. (2005)	Tìm kiếm trực tuyến, dữ liệu việc làm hàng tháng WordTracker, BLS và yêu cầu dữ liệu của Cục Lao động)	Mối quan hệ giữa các dữ liệu tìm kiếm trực tuyến và dữ liệu thất nghiệp chính thức / nguyên nhân trong thời gian dài liên quan đến giải thích sức mạnh thấp và hồi quy với nam giới trên 20 tuổi có ý nghĩa thống kê F.

4. Cơ hội cho các số liệu thông kê kinh tế và dữ liệu lớn là gì?

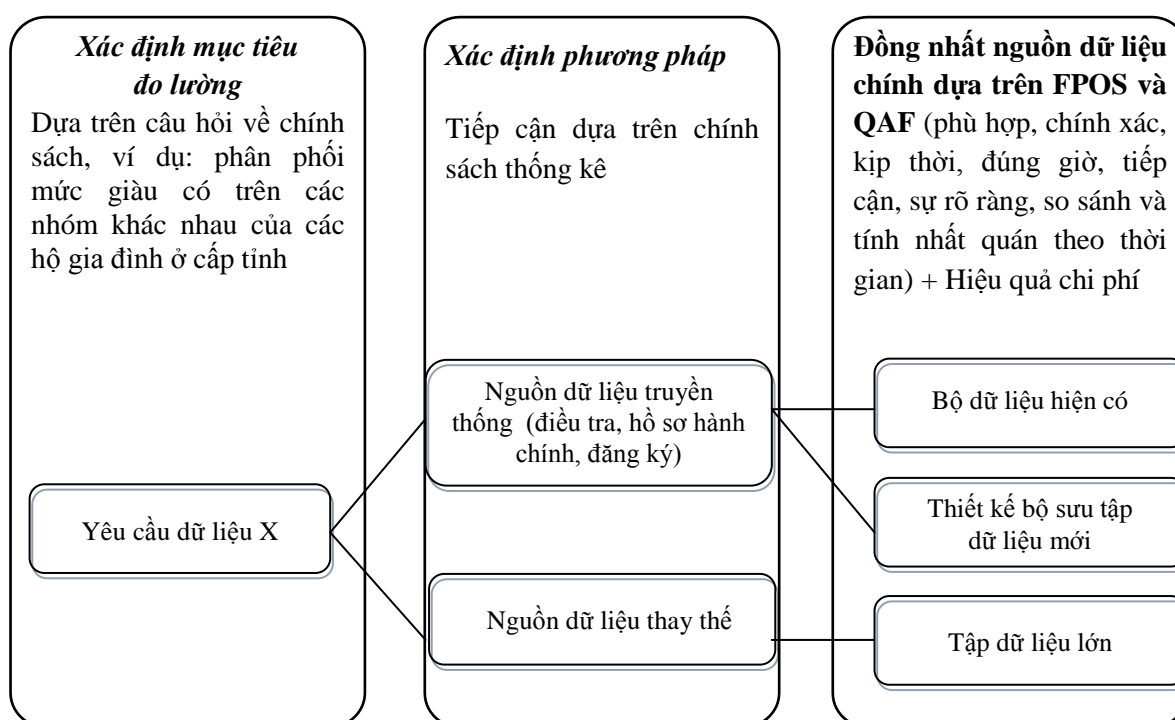
Tìm kiếm trực tuyến dữ liệu lớn là không có giới hạn điều này không có gì phải nghi ngờ. Tuy nhiên, các hoạt động trực tuyến diễn ra ngày càng nhiều, với việc tìm kiếm dữ liệu là miễn phí, nghiên cứu này đã sử dụng nhiều loại dữ liệu lớn để chứng minh sức mạnh của nó trong dự báo và với số liệu thống kê chính thức. Với dữ liệu chi tiết hơn, bài viết này lập luận đó là tiềm năng sử dụng dữ liệu tìm kiếm trực tuyến (và các loại dữ liệu lớn) để cung cấp số liệu về các khía cạnh phân phối thu nhập trong thống kê kinh tế vĩ mô. Điều quan trọng là phải tăng cường mối liên kết giữa các số liệu thông kê kinh tế và người dân mà hiện nay chưa được bảo đảm đầy đủ bởi số liệu thống kê kinh tế chính thức, ví dụ: về lâu dài, một người tiêu dùng là thành phần quan trọng trong bảng cân đối của hộ gia đình và hiểu về sản xuất hộ gia đình sử dụng theo phương thức riêng.

Có rất ít số liệu thống kê chính thức về khối lượng và giá trị thương mại điện tử nhưng dữ liệu lớn đã cho thấy xu hướng phát triển cả ở các nước phát triển cũng như các nước đang phát triển (www.e-marketer.com). Dữ liệu từ Liên hiệp Viễn thông Quốc tế (ITU) báo cáo về tình hình sử dụng internet phù hợp với điều này và cho thấy xu hướng ngày càng tăng trên toàn thế giới trong tiếp cận và sử dụng internet (xem biểu đồ tại phụ lục). Sự gia tăng này thậm chí còn rõ rệt hơn đối với các nước đang phát triển, nơi xuất phát điểm thấp, ví dụ như năm 2005 dưới 8% các cá nhân được sử dụng internet tại các quốc gia đang phát triển so với gần 1/3 trong năm 2014. Điều này cũng cho thấy một thực tế là mặc dù xu hướng tăng ở các nước đang phát triển nhưng không phải tất cả các phân tầng của xã hội sử dụng internet theo một cách thức giống nhau, ví dụ: tiên hành trong các hoạt động thương mại điện tử.

Qua công cụ phân tích Google Analytics, nguồn thông tin tiên tiến hơn so với GT, đưa ra chi tiết về cá nhân cho thấy lợi ích trong một trang web cụ thể, chủ đề hoặc sản phẩm cho các doanh nghiệp. Google Analytics có hồ sơ người dùng trong đó có thể cung cấp thông tin nhân khẩu học, công nghệ cho khách hàng sử dụng để đáp ứng được các nội dung của một trang web và các hành vi của các khách hàng. Các thông tin cá nhân được cung cấp bởi Google Analytics được thu thập thông qua các phương tiện khác nhau. Ví dụ, trong một máy tính cụ thể trình duyệt web sẽ được kết hợp với dữ liệu nhân

khẩu học hoặc có thể cung cấp các thông tin cá nhân có sẵn trên các trang mạng xã hội. Trong trường hợp của người tìm kiếm bằng cách sử dụng điện thoại di động, các ứng dụng trên điện thoại thông minh có thể cung cấp dữ liệu nhân khẩu học. Nhược điểm của nguồn dữ liệu này là chỉ có thể tìm được trên cơ sở riêng mỗi trang web cung cấp thông tin chứ không phải là một truy vấn tìm kiếm trực tuyến cụ thể. Ngoài ra, tùy thuộc vào sở thích của người sử dụng dữ liệu sẵn có sẽ được biến đổi ở mức cao.

Sơ đồ 1: Lựa chọn nguồn dữ liệu



Sơ đồ 1 gồm các bước lựa chọn nguồn dữ liệu cho bất kỳ yêu cầu dữ liệu dựa trên một câu hỏi về chính sách chung. Sự khác biệt trước khi có nguồn dữ liệu lớn là nguồn dữ liệu thay thế hiện nay được thêm vào như là một sự lựa chọn. Việc lựa chọn các

phương pháp hoặc các nguồn dữ liệu phụ thuộc trong ví dụ đầu tiên về chính sách thống kê của quốc gia. Các nước tiên tiến như Australia, có chiến lược dữ liệu lớn phủ rộng trong đó được xác định ưu tiên nghiên cứu dữ liệu lớn.

Mặt khác, việc lựa chọn các nguồn dữ liệu cụ thể, nên đưa vào các nguyên tắc cơ bản của thống kê chính thức và các yếu tố của khung đảm bảo chất lượng. Thông điệp chính của sơ đồ này là để cho thấy trình tự các bước đo lường không phải là khác nhau khi sử dụng dữ liệu lớn cho thống kê chính thức.

Trong sơ đồ 2 và 3 đưa ra các bước sử dụng nguồn dữ liệu lớn để cung cấp số liệu thống kê kinh tế về một khía cạnh phân phối thu nhập và thu hẹp khoảng trống dữ liệu tương ứng. Nguồn dữ liệu thay thế có thể hữu ích ở khía cạnh cung cấp số liệu phân phối thu nhập bởi khả năng bổ sung cho nguồn dữ liệu vi mô và bao gồm các dữ liệu dân số thường bị loại trừ trong trường hợp sau, ví dụ: các hộ gia đình giàu có (có sự tích lũy tài sản) và thậm chí cả các hộ thể chế (đặc biệt là nhà dưỡng lão, nơi mọi người có thể được dự kiến sẽ sở hữu nhất định / tài sản tài chính, phi tài chính). Như vậy để đánh giá này là cần thiết cho nghiên cứu các xu hướng các nhóm dân số tìm kiếm trong các loại dữ liệu lớn có sẵn, ví dụ: tìm kiếm trực tuyến hay dữ liệu thay thế.

Sơ đồ 2 gồm các bước về sử dụng dữ liệu lớn, ví dụ tìm kiếm trực tuyến cho một thành phần quan trọng của hàng tiêu dùng; phương tiện (được sử dụng bởi những người hoặc hộ cho lợi ích cá nhân của mình hay sự hài lòng). Như đã đề cập trước đó, hàng tiêu dùng bao gồm một phần của cải tích lũy của hộ gia đình và dịch vụ từ hàng tiêu dùng có tác động trực tiếp tới lợi ích của người dân, khi họ tìm kiếm trong sản xuất hộ gia đình sử dụng phương thức riêng. Điều này cũng được nhấn mạnh yêu cầu để ước lượng giá trị hàng tiêu dùng trong SNA 2008 (trong đó liệt kê hàng tiêu dùng như mục bản ghi) và trong thành phần sự tích lũy tài sản của hộ gia đình và bảng cân đối. Do đó, tăng cường thông tin về

hàng tiêu dùng đóng góp trực tiếp vào liên kết số liệu thống kê kinh tế với mọi người. Đây là một phần của quá trình được mô tả trong sơ đồ 3.

Một số chỉ tiêu mà chúng ta có thể xây dựng bằng cách sử dụng các nguồn dữ liệu thay thế có thể tham khảo các phương tiện sở hữu của các nhóm dân cư khác nhau ở các khu vực địa lý. Các cuộc điều tra hộ gia đình thường không đạt được mức độ mong muốn và mức độ phân tích chi tiết. Hồ sơ hành chính cũng có thể hội đủ điều kiện như dữ liệu lớn phụ thuộc vào khối lượng, tốc độ và các đặc điểm khác có thể cung cấp dữ liệu toàn diện trong phát triển hệ thống. Theo sơ đồ 2, giả sử các loại ưa thích của dữ liệu lớn là tìm kiếm trực tuyến từ những khối lượng từ khóa chính, bước đầu tiên là để chọn một từ khóa thích hợp sau đó được kiểm tra lại đối với các nguồn dữ liệu liên quan hiện có. Sau đó, các biến nhân khẩu học có sẵn - giả định rằng họ là chính xác, cần phải được kiểm tra đối với việc phân phối thu được từ số liệu thống kê chính thức, thông qua các dữ liệu tổng điều tra dân số hay các cuộc điều tra hộ gia đình. Ở giai đoạn phân tích, các thông tin phân phối thu được sau đó có thể được áp dụng để tổng hợp kinh tế vĩ mô như là những người có liên quan đến lĩnh vực hộ gia đình.

Trong trường hợp có được giá trị kinh tế của các phương tiện thuộc sở hữu trong một quốc gia, nguồn dữ liệu lớn như là nguồn dữ liệu bảo đảm sẽ cung cấp thông tin toàn diện hơn bao gồm cả sự cô về nhân khẩu học. Sơ đồ 3 mô tả các bước trong việc sử dụng dữ liệu lớn để thu hẹp những khoảng trống dữ liệu từ việc lựa chọn nguồn dữ liệu vào các giai đoạn phân tích. Trong trường hợp xác định giá trị, điều này đặc biệt quan trọng để phân biệt đầu hàng hóa cố định sử dụng trong hộ gia đình và đầu là tài sản cố định của doanh nghiệp hộ gia đình.

Sơ đồ 2. Sử dụng Bigdata đối với khía cạnh phân phối thu nhập, ví dụ quyền sở hữu phương tiện (xe)

<p>Lựa chọn bộ dữ liệu</p> <p>Ví dụ</p> <ul style="list-style-type: none"> • Từ khóa tìm kiếm trực tuyến, ví dụ như "bảo hiểm" và "sửa chữa / nhà để xe" cho xe ô tô, trang vàng dữ liệu cho các tìm kiếm địa chỉ doanh nghiệp • Kiểm tra các mối tương quan với bất kỳ số liệu thống kê chính thức hiện tại / nguồn dữ liệu khác, ví dụ thống kê dựa trên hồ sơ hành chính và các cuộc điều tra hộ gia đình 	<p>Chọn biến được phân tổ theo</p> <p>Ví dụ</p> <ul style="list-style-type: none"> • Vị trí, giới tính, tuổi tác, v.v... • Kiểm tra phân phối của các nhóm theo các đặc điểm nhân khẩu học • Số liệu tổng điều tra dân số và phân bố dân cư ở cấp quốc gia và địa phương • Dữ liệu thu nhập hộ gia đình và chi tiêu cho các mục trong câu hỏi, ví dụ: quyền sở hữu xe 	<p>Áp dụng trong phân tích</p> <p>Ví dụ</p> <p>Sử dụng phân phối của quyền sở hữu phương tiện thu được thông qua nguồn dữ liệu lớn tổng hợp kinh tế vĩ mô</p>
---	--	--

Sơ đồ 3. Sử dụng dữ liệu lớn để nâng cao tính sẵn sàng của dữ liệu, ví dụ ước tính phương tiện giao thông (xe)

<p>Lựa chọn bộ dữ liệu</p> <p>Ví dụ</p> <p>Giá trị của phương tiện sở hữu thông qua dữ liệu mua và sửa chữa, ví dụ: cơ sở dữ liệu bảo hiểm.</p>	<p>Xử lý dữ liệu</p> <p>Ví dụ</p> <ul style="list-style-type: none"> - Làm lên đến con số cấp quốc gia (nếu địa phương có thể làm) - Tính toán giá trị khấu hao (thường được dùng vào tài khoản trong dữ liệu bảo hiểm) - Phân biệt doanh nghiệp hộ gia đình 	<p>Áp dụng trong phân tích</p> <ul style="list-style-type: none"> - Trong xây dựng bảng cân đối - Mục Memo cho tài khoản quốc gia
--	--	--

Thách thức chính khi sử dụng các nguồn dữ liệu thay thế hoặc dữ liệu lớn để nâng cao kiến thức cơ bản điều đó có thể làm cho thông báo quyết định của chính phủ và các cá nhân trở lên tốt hơn là một phần việc, nó có xu hướng dịch chuyển trọng tâm từ việc lên kế hoạch các hoạt động thu thập dữ liệu với các thông số xử lý sẵn có linh hoạt hơn cho việc đối phó với điều đó xảy

ra với các dữ liệu để tồn tại. Trong đó, việc sử dụng dữ liệu lớn cho thống kê chính thức có thể làm thay đổi các quy trình kinh doanh hiện có, như kiểm soát nhiều hơn có thể xảy ra tác động đối với các hoạt động thiết kế, thu thập và phổ biến dữ liệu thô của các nhà thống kê chính thức. (Trong bối cảnh phát triển dữ liệu lớn, các mô hình kết hợp được coi như nguồn dữ liệu bổ sung đã xuất hiện. Ví dụ, công ty Premise thu thập dữ liệu các

mặt hàng thanh toán thông qua kỹ thuật nạo dữ liệu hoặc thông qua “điều tra viên” người sử dụng điện thoại thông minh để có thông tin giá cả của các mặt hàng được xác định trước và trong thời điểm bán hàng với mục đích xác định sự thay đổi giá cả. Mô hình kết hợp này là sự sáng tạo và cải cách trong việc thu thập dữ liệu. Hệ thống thống kê quốc gia cũng có thể áp dụng các phương thức thu thập dữ liệu như là một bước tiếp theo từ việc sử dụng máy tính bảng cho tới sử dụng điện thoại thông minh). Để giảm thiểu những điều không mong muốn cần có sự tham gia của các nhà hoạch định chính sách thống kê quốc gia và quốc tế trong việc xác định các lĩnh vực nghiên cứu cũng như các quy tắc ứng xử. Sự tham gia ngay từ đầu có khả năng kết hợp với kết quả của số liệu thống kê chính thức và nghiên cứu sẽ được chấp nhận và không bị đưa ra bàn luận.

Giữa các nguồn dữ liệu lớn có thể không phù hợp với nhau, vì vậy các nhà hoạch định chính sách kinh tế cần nói đến tính toàn diện (thông tin về các quốc gia / khu vực / nhóm đối tượng mà không được bao quát bởi kỹ thuật số, ví dụ những người không truy cập vào internet, v.v...) và so sánh các thông tin; ví dụ chuỗi thời gian cho các xu hướng và điều chỉnh thời vụ, bao lâu và mức độ nào thì từ khóa tìm kiếm trực tuyến có thể phục vụ được như là một máy chủ dịch vụ dữ liệu câu hỏi? Tất cả những thắc mắc này đòi hỏi tính minh bạch từ những người biên soạn dữ liệu lớn và phương pháp sử dụng trong xử lý dữ liệu. Nói cách khác, yêu cầu có siêu dữ liệu dễ hiểu và đầy đủ về dữ liệu lớn.

Một câu hỏi được đưa ra về sự riêng tư và bảo mật dữ liệu cá nhân. Với cải tiến mã điều khiển đối với sản xuất và phổ biến số liệu thống kê bao gồm số liệu chính thức và tư nhân. Thật thú vị, nghiên cứu được tiến hành bởi Nhóm chuyên gia cố vấn Boston cho thấy dữ liệu cá nhân như tên, sở thích, tuổi tác hay giới tính được coi như “không có gì cả” hoặc “xem nhẹ” bởi hầu hết người được hỏi không quan tâm tới đất nước hay tuổi tác trong Khảo sát toàn cầu về niềm tin tiêu dùng trong năm 2013 bao gồm 10.000 người tiêu dùng ở độ tuổi 18 trở lên tiến hành tại 20 nước phát triển và đang phát triển. Vị trí chính xác và thông tin về chồng, con là những dữ liệu nhạy cảm hơn cả thông tin thẻ tín dụng và dữ liệu tài chính.

“54% người tiêu dùng toàn cầu cho biết họ không cảm thấy phiền khi sử dụng các thông tin về họ nếu việc sử dụng không mang lại rắc rối, gây thiệt hại lợi ích, hoặc không làm hại họ” (Nhóm chuyên gia cố vấn Boston 2013). Do đó, đây là điều cần thiết để bảo đảm việc thu thập, lưu trữ và sử dụng các thông tin nhạy cảm để duy trì sự sẵn có của dữ liệu theo thời gian.

5. Kết luận

Thống kê chính thức được chuyển hướng tới một cách tiếp cận đầy đủ, điều này được khuyến khích và thúc đẩy bởi chương trình nghị sự phát triển sau năm 2015, yêu cầu gắn kết phương pháp luận và phân tích số liệu giữa các lĩnh vực thống kê khác nhau. Một cuộc thảo luận đã được diễn ra trong lĩnh vực thống kê kinh tế, đặc biệt là việc tìm kiếm các biện pháp thay thế về sự tiên bộ vượt xa hơn cả GDP.

Các nguồn dữ liệu đưa ra rất nhiều cơ hội để ứng phó với các chính sách cũng như những thách thức đối với việc nâng cao chất lượng số liệu thống kê chính thức. Trong bài viết này, các phần trước đây đã lập luận rằng, số liệu thống kê thu được từ các nguồn dữ liệu thay thế như dữ liệu lớn có thể sử dụng để bổ sung cho số liệu thống kê chính thức, các nhà hoạch định chính sách thống kê phải được tham gia ngay từ khâu đầu tiên. Bài viết này nghiên cứu tiềm năng các nguồn dữ liệu để tăng cường khía cạnh phân phối và thu hẹp khoảng cách số liệu trong nỗ lực thu hẹp số liệu thống kê kinh tế với mọi người, được hướng đến từ các phương pháp luận mới cũng như các nhu cầu đo lường mới nổi, bao gồm cả các chương trình phát triển quốc tế mới. Hàng tiêu dùng được chọn làm ví dụ vì đây là thành phần quan trọng trong Bảng cân đối kinh tế hộ gia đình, có thể cải thiện sự hiểu biết của chúng ta về phúc lợi của người dân khi chúng ta không có đủ dữ liệu.

Bài viết dựa trên các ví dụ sử dụng dữ liệu lớn (trên cơ sở dữ liệu tìm kiếm trực tuyến) trong thống kê kinh tế để thấy hầu hết các kết quả vẫn còn ở giai đoạn xác định mối tương quan giữa các số liệu thống kê chính thức hay dữ liệu hành chính có liên quan. Xu hướng ngày càng tăng việc sử dụng dữ liệu trực tuyến trong thống kê kinh tế. Bài viết làm sáng tỏ một số bước sơ bộ cho việc ra quyết định liên quan đến lựa chọn và áp dụng nguồn dữ liệu lớn tập trung vào hàng tiêu dùng (ví dụ các loại phương tiện) từ dữ liệu tìm kiếm trực tuyến.

Có một số thách thức về quy trình, kỹ thuật

và phương pháp luận xung quanh việc sử dụng dữ liệu lớn kết hợp với số liệu thống kê chính thức như việc thiếu một quy tắc ứng xử chung cho tất cả các bên liên quan (nhà nước và tư nhân), sự thiếu vắng siêu dữ liệu cho các nguồn dữ liệu lớn và các câu hỏi làm thế nào để đảm bảo sự riêng tư / bí mật của các nhà cung cấp dữ liệu cá nhân. Thách thức quan trọng nhất là duy trì sự tin tưởng của các cá nhân cũng là người trả lời trong các bộ sưu tập dữ liệu thống kê chính thức. Tuy nhiên, các tiềm năng của dữ liệu lớn chỉ có thể được khám phá nếu tính bền vững được đảm bảo, chắc chắn người cung cấp dữ liệu phải đáng tin cậy và chỉ sử dụng phù hợp mục đích phân tích. Nguồn dữ liệu lớn có tầm quan trọng đối với số liệu thống kê chính thức, được sử dụng phù hợp với các khuôn khổ đảm bảo chất lượng và không vi phạm các nguyên tắc thống nhất (ví dụ FPOS) bởi vì các tác động liên quan có thể gây tổn hại các bộ sưu tập dữ liệu truyền thống cũng được tính đến bởi hệ thống thống kê quốc gia.

Nghiên cứu tiếp theo liên quan đến dữ liệu lớn và thống kê kinh tế cũng như bất kỳ lĩnh vực thống kê nào, đòi hỏi việc thử nghiệm các phương pháp cũng như việc thu thập dữ liệu thực tế từ các nguồn dữ liệu thay thế để phân tích. Sử dụng dữ liệu lớn về hàng tiêu dùng chỉ là một khả năng trong việc nhấn mạnh những khía cạnh của phân phối thu nhập hộ gia đình trong thống kê kinh tế. Điểm mấu chốt là nghiên cứu chính sách có liên quan, trong đó khả năng có được đầy đủ thông tin của thống kê chính thức chứ không phải chỉ hạn chế để nghiên cứu xu hướng dữ liệu.

Tài liệu tham khảo:

- [1] E. Giovannini. "Understanding Economic Statistics. An OECD Perspective". OECD, Paris. 2008.
- [2] Eurostat. G20 Data Gaps Initiative (DGI)—background. Available at: [http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/G20_Data_Gaps_Initiative_\(DGI\)_%E2%80%933_background](http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/G20_Data_Gaps_Initiative_(DGI)_%E2%80%933_background)
- [3] OECD Guidelines for Micro Statistics on Household Wealth". OECD, 2013.
- [4] A. Leone. "The Balance Sheet Approach: Data Needs, Data at Hand, and Data Gaps". Statistics Department, International Monetary Fund. August 2009.
- [5] Stiglitz, Joseph and Amartya Sen and Jean-Paul Fitoussi. "Report by the Commission on the Measurement of Economic Performance and Social Progress". 2008. Available at http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf
- [6] W.D. Nordhaus and J. Tobin. "Is growth obsolete?" Economic Research: Retrospect and Prospect. Vol: 5 Economic Growth. 1972. Available at: <http://www.nber.org/chapters/c7620.pdf>
- [7] M. Fesseau et al. A cross-country comparison of household income, consumption and wealth between micro sources and national accounts aggregates. OECD Working Paper 62. OECD/Eurostat, 2013. Available at http://www.oecd-ilibrary.org/economics/a-cross-country-comparison-of-household-income-consumption-and-wealth-between-micro-sources-and-national-accounts_aggregates_5k3wdjrn7mv-en
- [8] M. Fesseau and M. Liviana Mattonetti. Distributional measures across household groups in a national accounts framework, Results from an experimental cross-country exercise on household income, consumption and saving. OECD Working Paper No. 53. OECD/Eurostat, 2013. Available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/gdp_and_beyond/documents/HH_WP1_Distributional_measures.pdf
- [9] OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth. OECD, 2013.
- [10] Open Working Group. Introduction to the Proposal of the Open Working Group for Sustainable Development Goals. July 2014. Available at: http://sustainabledevelopment.un.org/content/documents/4518SDGs_FINAL_Proposal%20of%20OWG_19%20July%20at%201320hrsver3.pdf
- [11] System of National Accounts 2008. United Nations, New York, 2009.
- [12] UN Global Pulse. Big Data for Development: Challenges and Opportunities. Global Pulse. New York, May 2012. Available at <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopmentUNGlobalPulseJune2012.pdf>

[13] The Philosophy of Data Science (series)-Rob Kitchin: The Big Data should complement small data, not replace them. 2014 Available at: <http://blogs.lse.ac.uk/impactofsocialsciences/2014/06/27/series-philosophy-of-data-science-rob-kitchin/>

[14] MIT Billion Prices Project website: <http://bpp.mit.edu/>

[15] N. Maclaren and R. Shanbhogue. Using internet search data as economic indicators. Bank of England. Quarterly Bulletin. 2011, Q2.

[16] L. Wu and E. Brynjolfsson. The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities. ICIS 2009 Proceedings. Paper 147. 2009.

[17] F. D'Amuri and J. Marcucci. Google it! Forecasting the US unemployment rate with a Google job search index. MPRA Paper, No. 18732, October 2009.

[18] N. Askatas and K. Zimmermann. Google Econometrics and Unemployment Forecasting. IZA DP No. 4201. June 2009.

[19] M. Ettredge et al. Using Web-based Search Data to Predict Macroeconomic Statistics. Communications of the ACM, Vol. 48, No. 11, November 2005, pp. 87-92.

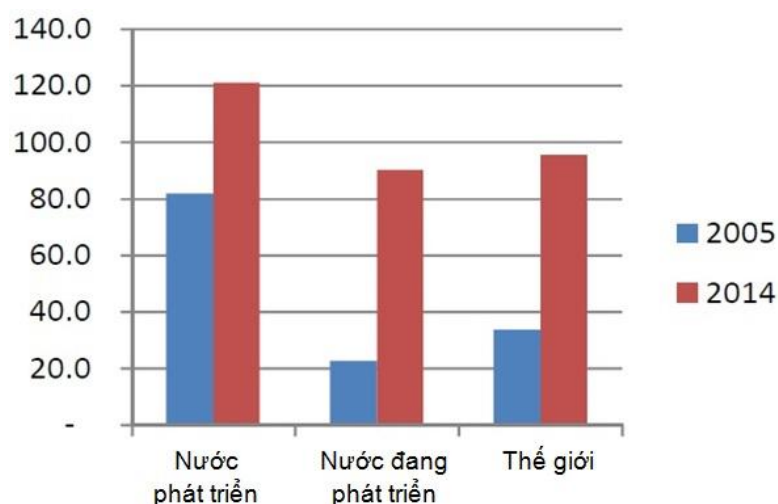
[20] Google Analytics website: https://support.google.com/analytics/answer/1012034?hl=en&ref_topic=1007027

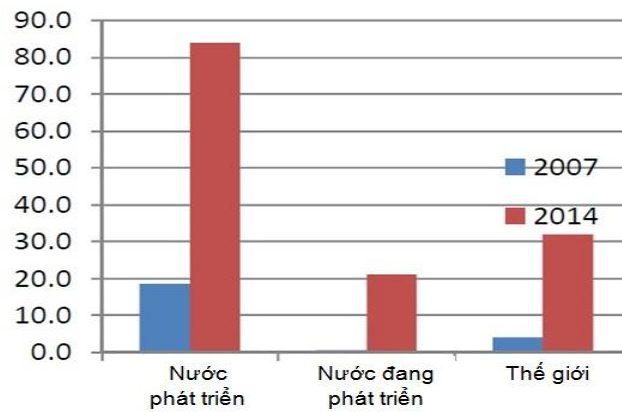
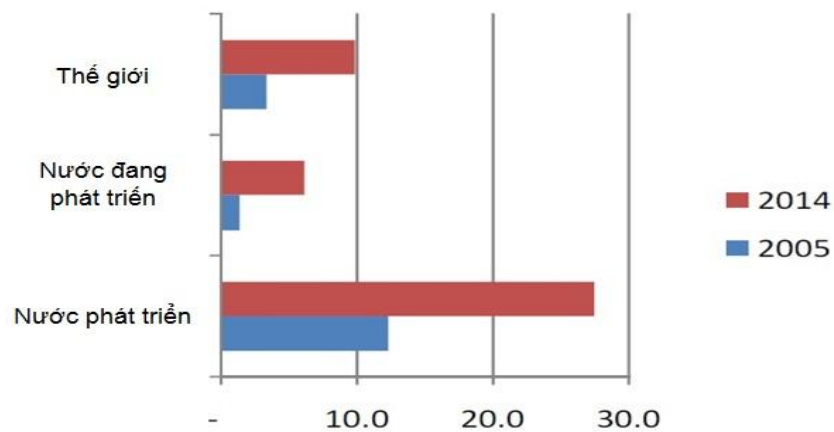
[21] The Trust Advantage. How to Win with Big Data. The Boston Consulting Group. November 2013.

[22] Premise website: www.premise.co

Phụ lục: Truy cập và sử dụng internet (Nguồn: ITU)

Biểu đồ a. Số thuê bao điện thoại di động (trên 100 dân)



Biểu đồ b. Hoạt động thuê bao di động (trên 100 dân)**Biểu đồ c.** Số thuê bao cố định (có dây) thuê bao di động (trên 100 dân)**Biểu đồ d.** Tỷ lệ sử dụng internet (%)