

MỘT SỐ KHÍA CẠNH KHI ĐÁNH GIÁ CHẤT LƯỢNG SỬ DỤNG DỮ LIỆU HÀNH CHÍNH TRONG THÔNG KÊ CHÍNH THỨC

Rudi Seljak, Cơ quan Thống kê quốc gia Cộng hòa Slovenia

1. Giới thiệu

Việc sử dụng dữ liệu hành chính trong sản xuất số liệu thống kê chính thức đang ngày càng trở nên phổ biến trong những năm gần đây. Mặc dù cần phải thấy rằng năm nay dữ liệu hành chính đã được sử dụng khá nhiều để sản xuất số liệu thống kê chính thức nhưng trên thực tế ý định sử dụng dữ liệu hành chính về cơ bản đã có thay đổi trong những thập kỷ qua. Nếu như trước đây dữ liệu hành chính chủ yếu được sử dụng nhằm mục đích xây dựng dàn mẫu hoặc các mục đích phụ liên quan đến sự thay đổi trong quá trình dự đoán thì giờ đây người ta ngày càng sử dụng nhiều dữ liệu hành chính để làm nguồn dữ liệu trực tiếp. Động cơ thúc đẩy thực hiện nhiều hoạt động như vậy đó là khả năng cắt giảm hầu như toàn bộ ngân sách cần thiết khi mà việc thu thập dữ liệu tốn kém trong trường hợp điều tra thống kê “cổ điển” được thay thế bằng việc thu thập dữ liệu từ hồ sơ hành chính rẻ hơn nhiều.

Các xu hướng giảm chi phí điều tra tất nhiên là không có gì mới và các nhà thống kê luôn phải đối mặt với yêu cầu cắt giảm chi phí. Tuy nhiên, chỉ có sự phát triển nhanh chóng môi trường công nghệ thông tin (CNTT) trong thập kỷ qua mà giờ đây cho phép xử lý nhanh chóng và hiệu quả các tập dữ liệu lớn mới có thể tạo nền tảng kỹ thuật làm cho việc sử dụng rộng rãi các nguồn dữ liệu khác nhau trở nên khả thi. Mặt khác, trình độ hiểu biết về lợi

ích của sự hợp tác chặt chẽ hơn giữa các cơ quan nhà nước khác nhau đã tăng lên đáng kể ở nhiều nước. Theo quan điểm của nhà sản xuất thống kê, kết quả hữu ích nhất của hợp tác sâu hơn như vậy là sẵn sàng chia sẻ dữ liệu. Mặc dù vẫn còn một số do dự chủ yếu được chứng minh thông qua các luận cứ về nguy cơ bị kiểm soát nhưng ngày càng có nhiều hồ sơ hành chính có sẵn cho các nhà sản xuất số liệu thống kê chính thức.

Mặc dù ưu điểm của việc sử dụng dữ liệu hành chính là khá rõ ràng, đặc biệt là trong lĩnh vực chi phí và giảm gánh nặng trả lời nhưng cũng cần thường xuyên xem xét các nhược điểm và thiếu sót có thể của biện pháp này. Trước hết, chúng ta phải chỉ ra việc chúng ta thường quá tin tưởng sẽ tìm thấy dữ liệu hành chính tại các cơ quan thống kê và rằng chúng không có bất kỳ lỗi nào mà quên rằng các cơ quan hành chính cũng sử dụng một số kiểu thu thập và cũng không tránh khỏi việc tạo ra các kiểu sai sót khác nhau trong dữ liệu thu thập được. Bên cạnh những sai sót về đo lường do sự “ẩn đi” của quá trình thu thập thì còn tồn tại một số khía cạnh khác về chất lượng mà rất cụ thể đối với các trường hợp điều tra dựa trên các dữ liệu hành chính và tất cả các khía cạnh cần được nghiên cứu càng chi tiết càng tốt.

Trong bài báo này, chúng tôi sẽ đưa ra cái nhìn tổng quan ngắn gọn (theo quan điểm

của chúng tôi) về các khía cạnh quan trọng nhất của quá trình đánh giá chất lượng trong trường hợp điều tra nơi sử dụng dữ liệu hành chính làm nguồn dữ liệu trực tiếp. Trong phần đầu của bài báo này, chúng tôi sẽ thảo luận một số vấn đề chung liên quan đến quá trình chuyển đổi từ phương pháp thu thập dữ liệu truyền thống sang sử dụng hồ sơ hành chính, cố gắng chỉ ra những thay đổi chính có được nhờ sự chuyển đổi này. Trong phần thứ hai, chúng tôi sẽ nêu cụ thể quy mô chất lượng theo quy định trong Hệ thống thống kê châu Âu. Chúng tôi sẽ thảo luận về vai trò khác nhau của quy mô này trong trường hợp sử dụng các phương pháp mới để thu thập dữ liệu và thảo luận về ảnh hưởng của những thay đổi trong quá trình đánh giá chất lượng đến việc báo cáo chất lượng. Cuộc thảo luận mang tính chất lý thuyết sau đó sẽ được bổ sung thêm hai trường hợp thực tế thu được tại văn phòng thống kê của chúng tôi.

2. Dữ liệu hành chính - một vài nhận xét chung

Trong phần này, chúng tôi thảo luận về ba khía cạnh xuất phát từ ba câu hỏi mà tất cả các câu hỏi này đều đề cập đến mức độ xem xét nói chung các tính năng chính của dữ liệu hành chính dự kiến sẽ được sử dụng trong quá trình thống kê.

2.1. Dữ liệu hành chính có phản ánh thực tế?

Khi sử dụng dữ liệu hành chính, chúng ta phải nhớ rằng dữ liệu ban đầu được thu thập không nhằm mục đích thống kê mà nhằm mục đích hành chính. Trên thực tế, các dữ liệu này thường được thu thập trong quá trình thực thi pháp luật hoặc thực hiện một số hành vi hành chính khác. Hậu quả trực tiếp của việc này là giới luật pháp phải mất nhiều lần mới xác định được dữ liệu hành chính. Nói

cách khác, dữ liệu hành chính có xu hướng chỉ cho thế giới biết rằng nó là "đúng luật" chứ không phải "đúng thực tế". Vì các nhà thống kê ít nhiều cố gắng mô tả cho thế giới biết rằng nó là "đúng thực tế" tuy nhiên sự khác biệt giữa "đúng luật" và "đúng thực tế" mô tả các hiện tượng quan sát được lại là một trong những điểm quan trọng khi xem xét, đánh giá chất lượng của dữ liệu hành chính trên cơ sở số liệu thống kê. Tất nhiên, không còn nghi ngờ rằng đã nhiều lần, hai giới đề cập ở trên là rất giống nhau và các dữ liệu hành chính có thể phục vụ hoàn hảo cho mục đích thống kê. Tuy nhiên, đôi khi không nên bỏ qua những khác biệt này.

2.2. Dữ liệu nào chính xác hơn - hành chính hay thống kê?

Đây là một câu hỏi rất chung chung thường được nêu khi đề cập đến chất lượng của một số nguồn dữ liệu hành chính. Khi xử lý nguồn dữ liệu cụ thể, có thể soạn các câu hỏi như: "Dữ liệu nếu thu thập theo cách cổ điển sử dụng bảng câu hỏi thống kê sẽ chính xác hơn các dữ liệu thu được từ các nguồn hành chính chứ?". Như thường lệ, các trường hợp câu hỏi quá chung chung sẽ không thể có câu trả lời chính xác. Việc dành ưu tiên cho một hay các nguồn dữ liệu khác chỉ đơn giản phụ thuộc vào quá nhiều yếu tố quá cụ thể và cần phải nghiên cứu riêng từng trường hợp. Ở đây, chúng tôi chỉ đưa ra một số đánh giá nói chung cần thực hiện khi xử lý các loại câu hỏi. Do vấn đề này đã được chỉ ra nên trước hết chúng ta phải ý thức được rằng cả hai loại dữ liệu này đều có thể có sai sót, tuy nhiên nguyên nhân dẫn đến sai sót thường khá là khác nhau. Nếu trong trường hợp điều tra thống kê cổ điển, chất lượng dữ liệu đầu vào phụ thuộc nhiều vào vai trò hữu ích của các công cụ đo lường (bảng câu hỏi, giấy hẹn, v.v) được sử dụng trong quá trình thu thập dữ liệu và phụ thuộc

vào trình độ và kinh nghiệm của người phỏng vấn thì độ chính xác của dữ liệu hành chính vi mô thường phụ thuộc vào các yếu tố đôi chút khác nhau. Hai trong các yếu tố này có thể được thể hiện thông qua các câu hỏi sau đây:

- *Hậu quả của việc báo cáo dữ liệu không chính xác là gì?*

Dưới đây, chúng tôi đề cập đến các dữ liệu cần cung cấp (bởi thể nhân hoặc pháp nhân) cho một số cơ quan hành chính và những nơi mà cơ quan này có quyền kiểm tra tính chính xác của dữ liệu được cung cấp và cũng có quyền phạt người trả lời nếu cung cấp dữ liệu không chính xác. Một ví dụ điển hình của cơ quan kiểu này là cơ quan thuế, đơn vị kiểm soát có hệ thống độ chính xác của dữ liệu đầu vào và có quyền phạt người được cho là báo cáo dữ liệu không chính xác. Do vậy, các dữ liệu về thuế thường được cho là có độ chính xác cao. Ở đây, chúng ta phải nhấn mạnh rằng chỉ có sự chính xác khi phân biệt những gì nên được báo cáo và những gì thực sự được báo cáo mới có nghĩa ở đây. Sự khác biệt về các khái niệm thu được là một câu chuyện hoàn toàn khác và sẽ được xử lý sau.

- *Lợi ích có thể thu được từ việc báo cáo số liệu không chính xác là gì?*

Việc báo cáo cho cơ quan hành chính số liệu không chính xác có thể mang lại một số lợi ích (vật chất hay phi vật chất). Do bản chất của con người là tìm kiếm các tiện ích cao nhất có thể cho bản thân và cộng đồng của mình nên cơ hội có được trong các trường hợp như thế này là "thích nghi" một chút với dữ liệu nếu sự thích nghi đó có thể mang lại lợi ích cao hơn. Kết quả có thể thu được trong các trường hợp này thường liên quan đến các câu hỏi về hình phạt xử lý nêu trên.

2.3. Có sự chặt chẽ chặt chẽ giữa dữ liệu hành chính của các nguồn khác nhau?

Câu hỏi thảo luận ở đây thường nảy sinh khi sử dụng dữ liệu hành chính làm nguồn phục vụ điều tra toàn diện ở đó số lượng lớn các dữ liệu cần được thu thập ở cấp vi mô. Cụ thể, trong những trường hợp này, chúng ta thường phải đối mặt với những tình huống là phải thu thập dữ liệu từ các nguồn dữ liệu khác nhau và điều này có thể gây nên mọi vấn đề khác nhau về tính toàn vẹn và nhất quán. Nếu tất cả các dữ liệu này được thu thập sử dụng phương pháp điều tra cổ điển thì tất cả những câu hỏi cần thiết sẽ được bao hàm trong bảng câu hỏi và chúng ta sẽ dễ dàng hơn nhiều để có được sự chặt chẽ giữa các dữ liệu. Mặt khác, dữ liệu từ các nguồn hành chính khác nhau có thể liên quan đến các giai đoạn khác nhau, sử dụng các đơn vị quan sát khác nhau, nhóm đối tượng khác nhau và có thể dựa trên cách tiếp cận khái niệm khác nhau. Nếu có một cuộc điều tra như vậy được thực hiện thì điều đặc biệt quan trọng sau đó là tất cả sự khác biệt này được nghiên cứu cẩn thận, ảnh hưởng của chúng đến chất lượng của kết quả sau cùng được giảm thiểu càng nhiều càng tốt thông qua quá trình thống kê và những thiếu sót có thể bắt nguồn từ phương pháp tiếp cận này được báo cáo minh bạch cho người sử dụng biết.

3. Dữ liệu hành chính - Tổng quan về các thành phần chất lượng

Bằng cách sử dụng khung đánh giá chất lượng đã được chấp thuận sử dụng rộng rãi trong Hệ thống thống kê châu Âu, chất lượng sản phẩm và dịch vụ thống kê được đánh giá thông qua sáu yếu tố về chất lượng: tính thích hợp, tính chính xác, tính kịp thời và tính đúng hạn, khả năng tiếp cận và tính rõ ràng, tính tương thích, tính chặt chẽ. Yếu tố thứ bảy, có tính bổ sung là *chi phí* và *gánh nặng* dựa trên sự chặt chẽ không phải là thành phần chất

lượng trực tiếp mà là một yếu tố quan trọng có thể ảnh hưởng đến tất cả các thành phần khác.

Nhìn vào định nghĩa các yếu tố chất lượng nêu trên, chúng ta có thể thấy các yếu tố này chịu ảnh hưởng rất khác nhau bởi một thực tế là quá trình thu thập dữ liệu cổ điển được thay thế bằng cách sử dụng dữ liệu hành chính. Nếu thành phần *khả năng tiếp cận và tính rõ ràng* một mặt không dễ dàng chịu ảnh hưởng bởi phương pháp thu thập dữ liệu thì *tính phù hợp* và *tính chính xác* cần được xử lý rất khác nhau trong những trường hợp này. Trong phần tiếp theo của bài viết, chúng tôi xem xét từng thành phần (trừ *khả năng tiếp cận và tính rõ ràng*) từ quan điểm điều chỉnh chúng cho các trường hợp sử dụng các nguồn dữ liệu hành chính hoặc kết hợp.

3.1. Tính phù hợp

Vai trò của tính phù hợp khi đánh giá chất lượng sẽ thay đổi đáng kể nếu sử dụng các nguồn hành chính. Trong trường hợp điều tra cổ điển, thành phần này gần như định hướng theo sản phẩm, chủ yếu đánh giá sự phù hợp của kết quả thống kê sau cùng ở mức độ đáp ứng nhu cầu của người sử dụng. Mặt khác, trong trường hợp sử dụng dữ liệu hành chính, tính phù hợp sẽ trở thành thành phần chủ yếu định hướng vào quy trình do phần lớn các yếu tố xác định tính phù hợp bắt nguồn trực tiếp từ phần đầu của quy trình khi các nguồn hành chính khác nhau được tập hợp lại để có thể sử dụng trong quy trình thống kê. Nói cách khác, nếu trong trường hợp đầu, tính phù hợp chủ yếu được nghiên cứu dựa trên quan điểm của người sử dụng thì trong trường hợp thứ hai, tính phù hợp phải trở thành một công cụ để đánh giá sự phù hợp của các nguồn đầu vào cho các mục đích đã lên kế hoạch.

Tính phù hợp gồm hai khía cạnh cần được nghiên cứu một cách đặc biệt triệt để trong giai đoạn xem xét sự phù hợp cho các mục đích của chúng ta:

- Các khái niệm về phương pháp xác định thay đổi trong các nguồn hành chính có đủ chặt chẽ với các khái niệm thống kê được nêu trong thiết kế điều tra của chúng tôi không? Thực tế là chất lượng trong trường hợp sử dụng dữ liệu hành chính chủ yếu được xác định thông qua sự khác biệt mang tính khái niệm như vậy.

- Thời gian tham khảo các thay đổi trong nguồn hành chính có phù hợp với thời gian hướng đến của cuộc điều tra không? Nếu có trường hợp nào không đúng, phải nêu rõ trong báo cáo.

3.2. Tính chính xác

Trong trường hợp điều tra cổ điển, tính chính xác thường được thể hiện thông qua các loại sai số khác nhau về mặt lý thuyết là các thành phần được xác định và mô tả chính xác nhất. Các thuật ngữ như sai số chọn mẫu, sai số do không trả lời, sai số đo lường,... được biết đến bởi tất cả những người thực hiện các cuộc điều tra thống kê. Tuy nhiên, rõ ràng là thiếu một khuôn khổ mạnh và nhất quán cho các trường hợp sử dụng nguồn dữ liệu hành chính hoặc kết hợp. Ở đây, chúng tôi chỉ đưa ra một vài suy nghĩ của mình về tính phù hợp cần nghiên cứu khi xem xét danh mục các sai số được nhiều người biết đến trong các trường hợp này.

- Trong hầu hết các trường hợp khi quyết định sử dụng dữ liệu hành chính, những dữ liệu này chủ yếu được xử lý nhằm mục đích điều tra và do vậy phương pháp lấy mẫu là không thích hợp trong các trường hợp này. Do đó, sai số chọn mẫu hiếm khi tồn tại trong trường hợp sử dụng dữ liệu hành chính. Tuy

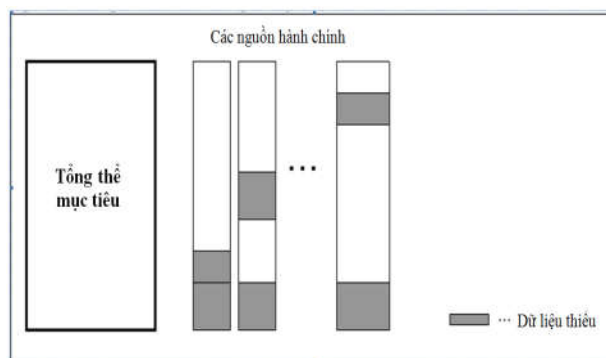
nhiên, nhiều khi cái giá của loại bỏ sai số chọn mẫu là sự gia tăng độ chệch trong các kết quả thu được của chúng tôi. Các nguồn chính của độ chệch là do nguồn hành chính không đầy đủ hay đôi khi liên quan đến ngày tham chiếu. Mặc dù đây không phải là nhiệm vụ dễ dàng nhưng cần nỗ lực để ít nhất là dự đoán sự gia tăng trực tiếp từ thực tế sử dụng các nguồn hành chính.

- Lỗi đo lường là một thành phần đặc biệt "rất khó đánh giá" trong trường hợp dữ liệu hành chính. Vì trong trường hợp này, quá trình thu thập được tách ra khỏi quá trình thống kê, chúng ta thường bị giới hạn bởi các quá trình chỉnh sửa nhằm tìm kiếm dữ liệu có sai sót hoặc đáng ngờ. Tuy nhiên, không thể xác minh các dữ liệu này tại nguồn dữ liệu. Trong các trường hợp như vậy, việc phối hợp chặt chẽ với các nhà cung cấp dữ liệu là rất quan trọng. Cụ thể, nhà cung cấp dữ liệu (cơ quan hành chính) có thể (ngoài các dữ liệu bản thân) đôi khi cung cấp thông tin hữu ích mà sau đó được sử dụng cho mục đích đánh giá chất lượng. Trong mọi trường hợp, chúng ta nên tránh việc coi dữ liệu thu được từ các nguồn hành chính là có chất lượng cao mà không cần bất kỳ chỉnh sửa nào.

- Khái niệm không trả lời có thể khá mơ hồ trong các trường hợp khi chỉ có dữ liệu hành chính được sử dụng trong các cuộc điều tra. Trong các trường hợp như vậy, thường khó phân biệt khái niệm này với khái niệm về sai số thu thập thông tin. Hãy để chúng tôi giả định đó là trường hợp "cổ điển" khi chúng tôi có danh sách các đơn vị được xác định từ trước thể hiện tổng thể mục tiêu của chúng tôi và sau đó lồng ghép một hay nhiều nguồn hành chính (sử dụng phương pháp liên kết trực tiếp hay gián tiếp) vào danh sách này. Điều chắc chắn không thể tránh khỏi sau quá trình lồng ghép là có những đơn vị đang thiếu một

số hoặc thậm chí là tất cả các thay đổi mục tiêu. Tình huống này chúng tôi trình bày trong hình sau:

Hình 1: Các giá trị bị thiếu sau giai đoạn lồng ghép



Vấn đề là ở chỗ trong tình huống như thế này, nhà cung cấp dữ liệu không cung cấp bất kỳ thông tin chi tiết nào nên rất khó để có thể nhận ra sự sai sót do không có câu trả lời về lý do dẫn đến sai sót do các vấn đề về phạm vi.

3.3. Tính kịp thời

Không cần phải tự thích ứng với khái niệm tính kịp thời khi chúng ta đang tiến tới sử dụng dữ liệu hành chính. Trong cả hai trường hợp, khái niệm này chỉ được định nghĩa đơn giản là khoảng thời gian giữa cuối giai đoạn tham chiếu và ngày cho ra kết quả. Tuy nhiên, việc thay đổi nguồn dữ liệu từ thống kê sang hành chính có thể có tác động khá khác biệt đến tính kịp thời. Câu hỏi đặt ra đương nhiên là việc sử dụng số liệu hành chính giúp cải thiện hay làm giảm giá trị của tính kịp thời. Tất cả những gì chúng tôi có thể nói ở đây là không có câu trả lời chung cho câu hỏi này. Tuy nhiên, theo kinh nghiệm của chúng tôi trường hợp suy giảm giá trị thứ hai phổ biến hơn nhiều so với trường hợp cải tiến trước đó. Trong hầu hết các trường hợp, chúng tôi phải đối mặt với sự đánh đổi cổ điển giữa cắt giảm chi phí cần thiết và kéo dài độ trễ thời gian hợp lý cho đến khi công bố kết quả.

3.4. Tính so sánh

Tính so sánh là một thành phần có thể chịu ảnh hưởng đáng kể khi chúng ta chuyển từ phương pháp thu thập dữ liệu cổ điển sang sử dụng số liệu hành chính đã thu thập được. Nếu chúng ta đang xem xét tính so sánh về địa lý thì rõ ràng là việc sử dụng các nguồn hành chính tại các nước khác nhau không thực sự giúp nâng cao trình độ so sánh. Cụ thể, các dữ liệu hành chính ở các nước khác nhau được xác định bởi các giới lập pháp khác nhau và để đạt đủ mức độ hài hòa sẽ khó hơn nhiều so với trường hợp khi tất cả các nước đồng ý sử dụng công cụ điều tra tương tự trong giai đoạn thu thập dữ liệu.

Quan điểm thứ hai về tính so sánh nên được xem xét trong quá trình đánh giá chất lượng là so sánh theo thời gian. Ngoài ra, ở đây chế độ thu thập dữ liệu có thể là một yếu tố quan trọng ảnh hưởng đến tính có thể so sánh. Dữ liệu “vấn đề” ở đây là những dữ liệu hành chính phụ thuộc rất nhiều vào luật pháp có xu hướng thay đổi thường xuyên và đáng kể. Trước hết, phải ghi nhận rằng cơ quan hành chính không quan tâm nhiều đến tính nhất quán của dữ liệu theo thời gian. Các tổ chức này quả thực lại chủ yếu hướng đến xác minh dữ liệu kiểu cắt ngang. Vì vậy, nghĩa vụ của tổ chức thống kê là sử dụng dữ liệu trong quá trình thống kê cẩn thận theo pháp luật cơ bản và thực hiện mọi hành động cần thiết để giảm thiểu các tác động có thể có của những thay đổi pháp luật đến kết quả thống kê.

3.5. Tính chặt chẽ

Nếu chúng ta xem xét tính chặt chẽ chỉ trong phạm vi chặt chẽ với kết quả thống kê từ các lĩnh vực khác (ví dụ như các tài khoản quốc gia) thì ảnh hưởng của việc sử dụng dữ liệu hành chính có thể vừa tích cực vừa tiêu cực. Trong trường hợp các cuộc điều tra thống

kê khác nhau “thực hiện” ở một khu vực như nhau (hoặc ít nhất là tương tự như nhau) sử dụng cùng một nguồn hành chính thì điều này sẽ làm tăng mức độ chặt chẽ của các kết quả. Mặt khác, nếu trong một cuộc điều tra có sử dụng nguồn hành chính và các nguồn thống kê khác thì ảnh hưởng đến sự chặt chẽ có thể là hoàn toàn ngược lại. Tuy nhiên, ngay cả trong trường hợp thứ hai thì thiếu sót này có thể biến thành lợi thế nếu chúng ta biết kết hợp một cách hợp lý các dữ liệu từ hai cuộc điều tra để nâng cao chất lượng. Ví dụ, trong cuộc điều tra cấu trúc sử dụng phương pháp điều tra toàn diện, có thể sử dụng các dữ liệu này để khắc phục một số khiếm khuyết sau cùng trong cuộc điều tra ngắn hạn dựa trên các dữ liệu hành chính.

4. Đánh giá chất lượng khi kết hợp dữ liệu thống kê và hành chính - hai trường hợp nghiên cứu

4.1. Các chỉ số doanh thu hàng tháng trong tổng mức bán lẻ và các dịch vụ khác

Chỉ số doanh thu hàng tháng là một trong những dự báo ngắn hạn quan trọng nhất mà văn phòng chúng tôi cung cấp. Kể từ năm 2006, tất cả các chỉ số này đã được tính toán trên cơ sở các dữ liệu thu thập thông qua các cuộc điều tra bưu chính “cổ điển” trong đó kiểu doanh nghiệp ngẫu nhiên được lựa chọn vào đầu mỗi năm và trong vòng 12 tháng sau đó dữ liệu ít nhiều được thu thập thành công từ các đơn vị này. Để giảm gánh nặng trả lời, đặc biệt là đối với các doanh nghiệp nhỏ, trong năm 2006 chúng tôi đã bắt đầu giới thiệu phương pháp nghiên cứu mới chủ yếu sử dụng dữ liệu về thuế GTGT hàng tháng thu được từ cơ quan thuế.

Phương pháp ước tính các chỉ số doanh thu hàng tháng mới này sử dụng hai loại dữ liệu. Đối với số lượng nhỏ các đơn vị lớn nhất

(theo doanh thu), dữ liệu vẫn được thu thập theo cách “cổ điển”, có nghĩa là các đơn vị được điều tra bằng bảng câu hỏi gửi qua bưu điện. Các đơn vị mà vẫn được điều tra theo phương pháp cổ điển chiếm 3% trong tổng số lượng đơn vị. Tuy nhiên, các đơn vị này lại chiếm hơn 50% tổng doanh thu. Đối với các đơn vị còn lại, phần lớn đơn vị, chúng tôi sử dụng dữ liệu của cơ quan thuế để ước tính doanh thu hàng tháng. Do vậy, Cơ quan Thống kê không liên hệ với các đơn vị này để phục vụ công tác điều tra. Như đã chỉ ra trong nghiên cứu khả thi được thực hiện trước khi đưa vào sử dụng phương pháp mới, các ước tính thu được từ các khoản mục trong biểu mẫu thuế không phải hoàn toàn phù hợp với định nghĩa mang tính phương pháp về doanh thu nhưng chúng chắc chắn có thể phục vụ cho các mục đích dự báo thay đổi (chỉ số).

Rõ ràng lợi ích lớn nhất của phương pháp mới này là giảm chi phí về phía văn phòng và giảm gánh nặng trả lời về phía các đơn vị báo cáo. Các thành phần chất lượng khác như thế nào? Chúng bị ảnh hưởng như thế nào bởi phương pháp mới? Ở đây, chúng tôi chỉ đưa ra một vài nhận xét về chủ đề này.

Tính phù hợp

Tính phù hợp của dữ liệu thuế GTGT được dùng để dự báo các chỉ số doanh thu hàng tháng đã được nghiên cứu cẩn thận thông qua các nghiên cứu khả thi nêu trên. Trên thực tế, doanh thu ước tính từ thuế GTGT không hoàn toàn được áp dụng để định nghĩa doanh thu về mặt thống kê chính thức. Tuy nhiên, các sai lệch này sẽ thực sự trở nên khó giải quyết nếu mức doanh thu là ước tính mục tiêu. Các kết quả của nghiên cứu khả thi cho thấy ảnh hưởng do định nghĩa không đúng về doanh thu là tương đối nhỏ hơn trong trường hợp khi chỉ ước tính các chỉ số. Tóm lại, có thể tóm lược rằng việc đưa vào sử dụng nguồn dữ

liệu mới đã làm giảm đôi chút tính phù hợp của nguồn dữ liệu, tuy nhiên lợi ích lớn hơn nhiều so với những sai sót.

Tính chính xác

Bên cạnh việc sử dụng nguồn dữ liệu mới, một sự thay đổi rất quan trọng khác đi kèm với phương pháp mới là sự chuyển đổi từ lấy mẫu ngẫu nhiên sang quá trình lựa chọn lấy mẫu giới hạn. Lý do chính của sự thay đổi này là khi sử dụng triệt để nguồn dữ liệu hành chính thì nhiều đơn vị lúc này sẽ có sẵn dữ liệu mà không phải mất bất kỳ chi phí nào và khá rõ ràng là việc sử dụng quy trình giới hạn phải dẫn đến nhiều kết quả chính xác hơn lấy mẫu ngẫu nhiên. Mặt khác, dữ liệu về thuế cũng không bao gồm tổng lợi nhuận. Điều này là do thực tế các đơn vị có doanh thu hàng năm nằm dưới ngưỡng nhất định không có nghĩa vụ phải báo cáo dữ liệu của mình. Ngoài ra, một số doanh nghiệp có nghĩa vụ báo cáo thì không bắt buộc phải báo cáo hàng tháng mà là hàng quý.

Hậu quả rõ ràng của sự thay đổi nêu trên trong quá trình lựa chọn tập hợp các đơn vị quan sát là sai số chọn mẫu đã được “thay thế” bằng sai số có hệ thống xảy ra khi thiết kế điều tra do phạm vi không đủ. Ưu điểm của phương pháp lấy mẫu là sai số chọn mẫu dễ dự đoán hơn nhiều (ít nhất là về mặt lý thuyết) so với sai số có hệ thống xảy ra khi thiết kế điều tra nhưng ít nhất là ở cấp độ hàng năm khi có nhiều thông tin phụ trợ hơn cho toàn bộ tổng thể có sẵn, sai số có hệ thống xảy ra khi thiết kế điều tra có thể ước tính khá đúng. Cho đến nay, các ước tính như vậy về sai số có hệ thống xảy ra khi thiết kế điều tra kết quả hàng năm cho thấy sai số này hầu như được dùng cho cấp thấp hơn hoặc ít nhất là cấp tương tự như các sai số chọn mẫu tương đối sử dụng phương pháp “cũ”.

Tính kịp thời

Tính kịp thời là một vấn đề quan trọng trong trường hợp thống kê ngắn hạn. Thời hạn nêu trong các quy định này là khá nghiêm ngặt và trong những năm gần đây, thời hạn này đã trở nên ngày càng ngắn hơn. Tại thời điểm này, thời hạn yêu cầu cao nhất là ở khu vực Tổng mức bán lẻ, nơi mà các kết quả đầu tiên phải được công bố trong vòng 30 ngày sau khi kết thúc tháng tham chiếu (T + 30). Các dữ liệu về thuế GTGT được "gửi" đến văn phòng của chúng tôi trong khoảng thời gian T + 45. Điều này có nghĩa là các dữ liệu này không phục vụ cho các dự toán thương mại bán lẻ đầu tiên. Do đó, quy trình sẽ như sau: Các ước tính thương mại bán lẻ đầu tiên hoàn toàn bắt nguồn từ các số liệu thống kê của các doanh nghiệp lớn, được thu thập từ các bảng câu hỏi gửi qua bưu điện. Do hầu hết các doanh nghiệp lớn này đã được điều tra nhiều năm nên không thể coi họ là người trả lời tốt, có dữ liệu được gửi đi một cách nhanh chóng và chủ yếu là có chất lượng rất cao. Sau này, khi thu được các dữ liệu thuế GTGT và dữ liệu từ các đối tượng trả lời sau cùng thì tất cả các dữ liệu này được lồng ghép vào và các dự đoán mới cho tổng mức bán lẻ và dự đoán đầu tiên cho các dịch vụ khác được tính toán và phổ biến.

Thực tế là khi xem xét tính kịp thời của số liệu thống kê ngắn hạn, cần đặc biệt lưu ý đến thỏa hiệp giữa thời gian và độ chính xác. Một lần nữa lại rất khó để có thể đánh giá phương pháp cũ và mới cái nào thành công hơn khi xử lý thỏa hiệp này. Tuy nhiên, theo ý kiến của chúng tôi (ý kiến có thể là rất chủ quan), sẽ rất khó để đảm bảo sự chính xác của số liệu trong một thời gian ngắn chỉ với dữ liệu điều tra.

Tính so sánh

Đảm bảo tính so sánh theo thời gian là

một nhiệm vụ rất khó khăn đối với dữ liệu thuế GTGT. Vấn đề xảy ra đối với các dữ liệu này là chúng chủ yếu được quy định bởi luật pháp và luật thuế có xu hướng thay đổi rất thường xuyên. Ngay từ đầu khi sử dụng dữ liệu thuế GTGT, chúng ta đã sử dụng bốn công thức khác nhau để tính doanh số ngoài các mục trong bảng câu hỏi về thuế GTGT. Những thay đổi thường xuyên sau đây về luật pháp và phân tích chính xác và cụ thể về ảnh hưởng của những thay đổi này đối với dữ liệu được cung cấp là rất quan trọng nếu chúng ta muốn duy trì đủ độ so sánh theo thời gian.

4.2. Điều tra về thu nhập và các điều kiện sống

Cuộc điều tra của châu Âu về thu nhập và điều kiện sống (EU-SILC) là dự án được thực hiện nhằm mục đích xây dựng cuộc điều tra phù hợp của châu Âu để thu thập số liệu thống kê so sánh về phân phối thu nhập và loại trừ xã hội từ các nước thành viên EU, Na Uy và Ai-xơ-len. Dự án đã được triển khai vào năm 2003 (lúc đó vẫn còn dựa trên thoả thuận) tại 6 nước thành viên châu Âu, đã được mở rộng sang 12 nước thành viên "cũ" vào năm 2004, Estonia và Ai-xơ-len và sau đó vào năm 2005 bao gồm tất cả các nước thành viên, Na Uy và Ai-xơ-len (vào thời điểm đó).

Tại Slovenia, EU-SILC lần đầu tiên được thực hiện vào năm 2005. Trong giai đoạn lên kế hoạch và xây dựng cuộc điều tra, chúng tôi đã cố gắng làm theo đề nghị của Eurostat là nên sử dụng càng nhiều nguồn dữ liệu có sẵn càng tốt để giảm gánh nặng trả lời và kết quả là tăng tỷ lệ phản hồi. Do đó, chúng tôi đã cẩn thận nghiên cứu tất cả các nguồn hành chính hiện có và chất lượng của chúng để phân bổ tất cả các nguồn có thể làm nguồn dữ liệu phục vụ công tác điều tra. Do vậy, ở Slovenia dữ liệu vi mô dùng cho EU-SILC được thu thập

từ ba loại nguồn. Phần dữ liệu đầu tiên được thu thập bằng cách thực hiện cuộc điều tra “cổ điển” sử dụng phương pháp CAPI và CATI, phần thứ hai được thu thập từ các nguồn thống kê khác và phần thứ ba được thu thập từ sổ đăng ký và các nguồn hành chính. Trong số các nguồn còn lại, tất cả các thay đổi liên quan đến thu nhập (mà thường được coi là thay đổi rất nhạy cảm) được thu thập từ các nguồn hành chính khác nhau.

Mặc dù sử dụng triệt để các nguồn hành chính mang lại nhiều lợi ích, đặc biệt là trong lĩnh vực gánh nặng trả lời và giảm chi phí điều tra nhưng cách tiếp cận như vậy cũng có thể gây nên những bất lợi nhất định. Trong phần này, chúng tôi sẽ thảo luận ngắn gọn về một số khía cạnh chất lượng của cuộc điều tra này liên quan đến sử dụng dữ liệu hành chính.

Tính phù hợp

Khi thiết lập cuộc điều tra, tất cả các nguồn cần sử dụng trong cuộc điều tra này đã được nghiên cứu cẩn thận và chỉ những người có liên quan đã được lựa chọn mới được sử dụng thường xuyên trong quá trình sản xuất dữ liệu. Do đó, theo quan điểm cơ bản mang tính phương pháp, tất cả các nguồn đều đáp ứng các yêu cầu thống kê. Tất nhiên là vẫn còn một số sai lệch bắt nguồn từ bản chất của dữ liệu hành chính. Tuy nhiên, theo kết quả phân tích, các ảnh hưởng này phải là rất quan trọng. Có lẽ vấn đề lớn nhất ở đây là một phần dữ liệu (dữ liệu liên quan đến thu nhập) có thời gian tham chiếu khác so với tất cả các dữ liệu điều tra. Đây là hậu quả trực tiếp của việc sử dụng các dữ liệu về thuế và không thể khắc phục trong quy trình thống kê. Tuy nhiên, những khác biệt này được nêu rõ trong các phần giải thích khi phổ biến kết quả.

Tính chính xác

Tính chính xác bản thân nó là một khái

niệm đa chiều. Do đó, có nhiều khía cạnh về tính chính xác có thể nghiên cứu, đặc biệt trong các cuộc điều tra phức tạp chẳng hạn như EU-SILC. Hệ quả tích cực nhất khi sử dụng dữ liệu hành chính để cho kết quả chính xác là thực tế rằng các câu hỏi ngắn hơn nhiều và tất cả các câu hỏi nhạy cảm được bỏ qua. Điều này chắc chắn sẽ dẫn đến tỉ lệ đáp ứng cao (đơn vị và hạng mục) và hy vọng là có ít sai số về đo lường. Nếu chúng ta đang xem xét các mục dữ liệu cụ thể thì có thể thảo luận xem nguồn nào (hành chính hoặc thống kê) có thể cho dữ liệu chính xác hơn trong từng mục. Đối với dữ liệu liên quan đến thu nhập, chúng tôi tin rằng nguồn hành chính tốt hơn nguồn thống kê. Tuy nhiên, trong một số trường hợp khác, ví dụ nơi sử dụng nhiều nguồn hành chính cho một sự thay đổi (ví dụ tình trạng việc làm) thì tình hình có thể diễn ra ngược lại.

Tính kịp thời

Các mốc thời gian của kết quả là khía cạnh chất lượng mà chúng ta mất nhiều nhất do sử dụng dữ liệu hành chính. Vấn đề chính ở đây là các dữ liệu thu nhập này do cơ quan thuế cung cấp. Tất nhiên là cơ quan thuế cần có thời gian để thu thập dữ liệu tổng thể, cần có quy trình và cần xác minh chúng, sau đó giao chúng cho văn phòng của chúng tôi, tại đây chúng được lồng ghép vào quy trình thống kê. Chúng tôi dự đoán rằng việc phát hành chậm do thu thập dữ liệu thuế chậm là khoảng 10 tháng. Tuy nhiên, tất cả các ưu điểm khi sử dụng dữ liệu hành chính vẫn còn vượt xa so với thiếu sót này.

5. Kết luận

Trong bài báo này, chúng tôi đã cố gắng làm sáng tỏ một số khía cạnh khi đánh giá chất lượng của cuộc điều tra sử dụng dữ liệu hành chính làm nguồn dữ liệu trực tiếp. Do nội dung

này đã được nêu trong nhiều bài báo khác nên khuôn khổ đánh giá chất lượng được triển khai sử dụng trong hệ thống thống kê châu Âu chủ yếu được thiết kế dành riêng cho các cuộc điều tra “cổ điển” ở đó dữ liệu được thu thập bằng cách sử dụng bảng câu hỏi thống kê. Trong những năm gần đây, do mức độ sử dụng dữ

liệu hành chính gia tăng rất nhanh nên khái niệm về nhu cầu đánh giá chất lượng đã điều chỉnh ngày càng rõ ràng hơn. Trong bài báo này, chúng tôi đã cố gắng sử dụng lý thuyết tổng hợp và hai ví dụ cụ thể, thực tế của Cơ quan Thống kê Slovenia để có những đóng góp khiêm tốn cho chủ đề phức tạp này.

Tài liệu tham khảo:

1. Lyberg L. et al.: *Khuôn khổ và chất lượng điều tra*, Wiley, 1997.
2. Seljak R., Flander Oblak A.: *Đánh giá chất lượng tổng điều tra của Slovenia dựa trên số đăng ký năm 2011*; Bài trình bày tại Cuộc họp giữa UNECE và Eurostat về Tổng điều tra Dân số và Nhà ở tại Geneva từ ngày 13 đến 15 tháng 5 năm 2008.
3. Seljak R., Ostrež T.: *Báo cáo chất lượng tại SORS – Trải nghiệm và Viễn cảnh tương lai*. Bài trình bày tại Hội nghị châu Âu về Chất lượng và Phương pháp Thống kê chính thức tại Helsinki, Phần Lan từ ngày 4 đến ngày 6 tháng 6 năm 2010
4. Seljak R., Zaletel M., “*Dữ liệu thuế - Phương tiện cần thiết để giảm gánh nặng trả lời các cuộc điều tra ngắn hạn*”, Bài trình bày tại Hội nghị quốc tế về Điều tra thiết lập, Montreal 2007
5. Wallgren A., Wallgren B.: *Thống kê dựa trên sổ sách; Dữ liệu hành chính dùng cho các mục đích thống kê*: John Wiley & sons, 2007.
6. Nhóm công tác “*Đánh giá chất lượng thống kê*”: *Định nghĩa chất lượng thống kê*. Tài liệu về phương pháp, cuộc họp lần thứ 6 tại Luxembourg được tổ chức vào ngày 2 và 3 tháng 10 năm 2003.

Thu Hiền, Nhật Linh (dịch)

