UNDERSTANDING CFS TRADE-OFFS

2020 CFS WORKSHOP BREAKOUT GROUP BACKGROUND INFORMATION

In our breakout sessions, participants will discuss the trade-offs we are willing to make between mode/geography/commodity detail to provide BTS/Census with greater understanding of user needs. This overview is designed to provide a better understanding of the framework within which these trade-offs are bounded so that our discussions can start from a shared understanding. This will focus describing:

* How the privacy/disclosure rules work and are applied
* How the tradeoff between geographic, commodity, and mode detail works
* What the finest and mode detail is in the data
* Whether there are less-detailed commodity aggregations than 2-digit SCTGs that could be accepted in exchange for greater geographic or modal detail

**Nondisclosure**

The “rule of three” comes into play quite a bit, although there are ways to overcome it sometimes (and publish data). There must be at a minimum three reporting firms in any breakdown, such as a certain commodity in a certain metro area. If there were only one it would often be easy to tell who’s reporting; if there were two it’s possible for each reporter to deduct what the other is doing. With three or more, numbers can be reported. Further, there are masking procedures that function to anonymize who reported, which are very useful in the microdata file.

Number of shipments comes into play as well. When there are very few shipments (fewer than five, as an example), then estimates might not get made. Once there are ten or fifteen, the data are likely sufficient. When there are 30 to 50 or more, we can have high confidence in the data.

While the number of establishments reporting is a critical threshold to meet, it is more often the number of shipments that creates non-disclosure issues.

**Details**

There is origin ZIP and destination ZIP for geography. Nondisclosure is a big issue for origin ZIP data because it becomes less plausible to conceal the establishment’s identity. Destination absent origin is likely more viable if there is a high volume of inbound-shipment data, but then only for certain commodities (i.e., small packages headed to residences). Tabulations below the county level can pass nondisclosure threshholds, although only for the largest counties. For the smallest counties – which are actually most of the some 3,300 in the country – it becomes challenging, even when commodities get rolled up.

That said, there may be methods to create sub-county geographies to allow analysis of other the data. The map below is an example of one such method.



*Potential for thinking differently about geography*

Modal data starts with what the respondents provide. If the shipment can’t credibly take the reported sequence, then the mode is imputed. There are checks to make sure that shipments that fall beyond the typical use of the reported conveyance get scrutinized. There are some missing details in the imputation of parcel model, especially for rail. The challenge is that shippers don’t always have enough of that information.

The figure below describes the current hierarchy and categories for collecting and reporting modal data. Other approaches are possible. With changes in the economy and supply chains, such as the continued growth of e-commerce, new categories may better describe what is taking place. Sometimes “multiple modes” does not provide the detail that is needed, so there may be improvements that can be made to collect and report on those.



*Would different modal categories be more helpful?*

**Commodity Aggregations**

Historically, commodity detail is collected at a detailed level, but reported on a more aggregated level collapsing the commodities into 9 groups of 2-digit SCTG codes. That hierarch can further collapse into four larger 2-digit groupings or to no commodity detail at all. This framework is shown below. One of the challenges with this structure is that it groups commodities that may not share transportation characteristics. There are likely methods to create categories to allow for the roll up of goods with similar transportation characteristics. There is more flexibility to modify the commodity framework than either geography or mode.

Moreover, in 2022 the plan is to no longer ask respondents for commodity codes and instead derive the classifications based off the reported descriptions. The SCTG scheme will remain because it is mature and aligns with the existing classification algorithms, but it is possible to move to another system.



*The commodity framework has greater flexibility than geography or mode*