From Transactions Data to Economic Statistics:
Constructing Real-time, High-frequency, Geographic Measures of Consumer Spending

Aditya Aladangady, Shifrah Aron-Dine, Wendy Dunn,
Laura Feiveson, Paul Lengermann, and Claudia Sahm[1]

Question: Accurate information about consumer spending is crucial for monitoring macroeconomic conditions and evaluating the welfare effects of policies and events. However, consumption can be difficult to measure in practice, particularly in a timely and geographically detailed manner. For instance, official estimates of retail sales, constructed by the Census Bureau from surveys of retail establishments, are available only at the monthly frequency after a lag of about two weeks, and are subject to considerable revision. These Census estimates also do not include any subnational detail, so for regional analyses, researchers and policy-makers must rely on other data sources; most often, however, these data sources have limited sample sizes at smaller geographies and are only available after a substantial lag.

Methods: To help fill these information gaps, we transform a raw dataset of card transactions into indexes of consumer spending that are timely, high frequency, and geographically detailed. The transactions are from First Data, a global payment technology company that processes roughly $2 trillion worth of card transaction volumes per year. With technical support from Palantir, we filter, aggregate, and fully anonymize the set of card transactions. This methodology consists of three main steps: the selection of a subset of "well-behaved" merchants from the payment processor's full set of merchant clients; the construction of nominal growth rates representative of U.S. geographies and industries and benchmarked to the 2012 Economic Census levels; and the seasonal adjustment of the daily data.

Data: Our approach yields a time series of daily retail spending, from 2010 to the present, for a number of industry categories, and various geographies (nation, U.S. states, and most large metropolitan statistical areas). Our new spending measures are conceptually similar to those reported in the Census Bureau's Monthly Retail Trade Survey (MRTS), the source data used to construct roughly one-third of consumer spending published in the national accounts. We validate our data construction procedures by comparing our new spending indexes with official statistics at the national level and find that the measures are highly correlated despite their distinct data sources and construction methods.

Preliminary Results: In an application of these new data, we use our spending indexes to track and quantify the spending effects of Hurricanes Harvey and Irma. This application extends the methods in our earlier research note (2016). We conclude that Hurricanes Harvey and Irma caused severe yet temporary disruptions to retail sales in Texas and Florida, and that together the two storms reduced national GDP growth by almost ½ percentage point (annual rate) in the third quarter of 2017. On balance, these data suggest that little of the reduced spending associated with Hurricanes Harvey and Irma was offset by higher spending in the days before or just after the storms. For this hurricane study, our dataset was more useful than the official Census retail sales release, which is limited by its national sampling frame and its monthly—as opposed to daily—frequency. More broadly, our new data set makes it possible to study other short-lived, geographically concentrated events or shocks. For other examples, see our research notes of sales tax holidays (2017) and the delay in the earned income tax credit (2018).

[1] We thank Aaron Jaffe, Dan Moulton, and other staff at Palantir and First Data for their many contributions to this project. The views expressed here are those of the authors and not necessarily those of other members of the Federal Reserve System.