Addendum to Intuit QuickBooks Small Business Index

Ufuk Akcigit Raman Chhina Seyit Cilasun Javier Miranda

Eren Ocakverdi Nicolas Serrano-Velarde

September 5, 2024

New revenue series for the US focuses on small business revenue growth

Since its launch in March 2023, the Intuit QuickBooks Small Business Index has provided high-quality monthly jobs reports measuring small business employment in the US and Canada and small business job vacancies in the UK.[1] In June 2024, the Small Business Index was expanded in the UK to include employment as well as job vacancies.[2] As described below, we now expand the methodology of the Small Business Index again to include a new small business revenue series in the US. The new series uses data provided by a sample of QuickBooks Payroll customers who use QuickBooks Online to manage their business finances. Using official statistics, this data is reweighted to mirror the US population of small businesses rather than Intuit's customers. The revenue series provides information that is independent of that contained in the employment series. Firm revenue responds in advance of the employment series and can provide insights into the relationship between employment and revenue in response to business cycles and economic shocks broadly.

How QuickBooks tracks revenue data

The QuickBooks platform helps subscribers automatically track various types of transactions in and out of their savings and checking accounts including bank deposits, credit card transactions, payments received, bills paid, and expenses in real time. QuickBooks automatically records money-in transactions in the bank feed as bank deposits. Users can classify transactions as revenue by using bank rules or by transaction individually. QuickBooks provides an initial categorization by matching transactions to sales receipts and invoices on the platform, but users are required to review and confirm that income transactions are correctly classified as revenue. Users typically do this with a 3 to 6 month lag — often when they need to meet their income tax reporting obligations. We use these revenue-classified monthly transactions as the basic input into the new revenue series of the Small Business Index. Revenue-classified deposits are enhanced with predicted revenue transactions for the most recent months. The resulting series are aggregated at the firm-month level before we produce the final revenue series of the Small Business Index.

---

[1] In the US, the employment series relies on official statistics from the U.S. Bureau of Labor Statistics (Job Openings and Labor Turnover Survey and Business Employment Dynamics) for firms with 1 to 9 employees to both benchmark and discipline the QuickBooks Payroll sample data — including firm entry and exit, as well as firm expansion and contraction. In Canada, for the same purpose, the employment series relies on official statistics from Statistics Canada for firms with 1 to 19 employees. Similarly, in the UK, the job vacancy series relies on official statistics from the Office for National Statistics for firms with 1-9 employees. For details, see Akcigit et al. 2023. Intuit QuickBooks Small Business Index: A New Employment Series for the US, Canada, and the UK.

[2] For details, see June 2024 Addendum to the Intuit QuickBooks Small Business Index methodology.

We develop a simple model to predict revenue transactions at the firm level based on the overall level of bank deposits. To this end, we first define a net deposit amount at the firm level as total deposits, minus loan disbursals, in the firm's savings and checking accounts. We then run the following specification which we use to predict revenue out of sample for firms with incompletely classified transactions:

$$R_{it} = \beta_0 + \beta_1 ND_{it} + \beta\ X\ + e_{it}$$

where R_{it} is revenue for firm i in month t, ND_{it} is net deposits for firm i in month t, $\boldsymbol{X}$ is a vector of month by sector and state dummies, and e_{it} is the error term which we assume to be uncorrelated. The predicted revenue is then deflated using the CPI. We create a missing industry dummy whenever the industry of the firm is missing from the data. The fit of the regression is high with an r^2 of 0.57.

Official statistics

National statistical offices such as the U.S. Bureau of Labor Statistics periodically release employment statistics as part of their mandate to provide policy-relevant information about the economy.[3] This is, however, not the case when it comes to business revenue data. As such, our use of official statistics in the new revenue series is limited to, 1) reweighting the data to ensure the population of QuickBooks customers in the US mirrors the true US population of business, and 2) benchmarking at low frequencies when this information is available. Reweighting is done using official business population numbers of business by size, region, sector, and year. To benchmark the data provided by QuickBooks customers, we make use of the U.S. Census Bureau's Statistics of US Businesses (SUSB) revenue numbers available from the economic census that is produced every five years. This is the only official source for data on the total revenue for enterprises of different sizes in the US.[4] An important strength of the SUSB data for our purposes is that it provides estimates of the total revenue for private sector businesses by size class and industry.

A modified methodology

The methodology we use to produce the employment series of the Small Business Index is designed to address three challenges when using data provided by QuickBooks customers:

1. The large growth in the use of the QuickBooks platform which generates drift in the growth rates from the large number of new users
2. The large number of spurious entry and exit events from entry and exit into the platform
3. Potential representation issues from the select set of users that choose to use the QuickBooks platform.[5]

As noted above, we are not able to discipline the revenue data against high-frequency official revenue statistics since these are not available. Instead, we impose two restrictions to limit the impact of the

[3] NSO produce monthly frequency economy-wide estimates, and more detailed low frequency administrative records based statistics with a lag. The more detailed series include statistics for small business.
[4] See https://www.census.gov/data/tables/2021/econ/susb/2021-susb-annual.html
[5] We document and address these in detail in Akcigit et al. (2023).

QuickBooks platform growth and entry/exit volatility on the revenue series. First, we limit revenue data to include only continuer firms with at least 3 months of revenue data. By doing so we purge all entry and exit events and any QuickBooks customers who use the platform for a very short duration. Second, we remove revenue outliers; that is, firms with monthly revenue larger than the 98th percentile of the revenue distribution and those reporting negative revenue.[6]

We proceed by producing seasonally-adjusted gross revenue expansions and gross contractions using the X-13-ARIMA SEATS program for each region and sector in the US sample of QuickBooks customer data. For this, we create region and sector cells as defined in the QuickBooks customer data. We exclude entry and exit from this series since this is highly volatile in the US sample and leads to noisy results. We compute net growth rates for each cell using the seasonally adjusted series. The additive properties of each series are maintained by adding the seasonally adjusted expansion and contraction components when creating the net growth series.

Our intent is to construct a nationally representative net growth series for small businesses in the US by constructing a series using appropriate industry times region cells weights. To construct the weights, we calculate the employment shares for cells including 50 regions, 12 sectors, and year interactions in the QBO sample and QCEW (official data).[7] We then calculate region-sector weights as the ratio of the QCEW to QBO shares (S) of establishments (E) as follows:

$$W^{srt} = \frac{S^{srt}_{QCEW}}{S^{srt}_{QBO}} = \frac{E^{srt}_{QCEW}/E_{QCEW}}{E^{srt}_{QBO}/E_{QBO}},$$

where *s*, *r*, and *t*, denote sector, region, and year. The QCEW are produced annually so we update the weights at the same frequency. Ideally, we would like to update the weights at higher frequencies, but since this is the only data available and the shares of activity don't change dramatically from year to year we proceed with this approach.

The QCEW tabulations include detailed size (including 1-9 employees), by industry and region cells. Some of the cells in QuickBooks customer data are scarcely populated, leading to potentially noisy estimates in some industry times region combinations. To avoid this, we replace low-sample region-sector combinations (cells with fewer than 30 observations in the QuickBooks customer data using the following 2-step procedure:

4. If the region-sector has a sample including at least 30 observations a given year but not for previous years, then we replace the weight from those low-sample size years with the latest year that has >=30 observations.

5. Otherwise, replace the region-sector weight with the overall national sector mean. Note we thus assume that the industry distribution follows the national average.[8]

---

[6] The resulting series is well behaved but it is by construction not reflective of revenue growth from firm entry and exit dynamics. We caution users of this limitation because firm entry and exit dynamics can accentuate incumbent dynamics.

[7] The 50 US states, excluding US territories and the District of Columbia. 7 NAICS sectors, including: sector 22 "Utilities", sector 23 "Construction", sector 31-33 "Manufacturing", sector 42 "Wholesale Trade", sector 44-45 "Retail Trade", sector 48-49 "Transportation and Warehousing", sector 51 "Information"; and 5 NAICS supersectors, including: sectors 11 and 21 "Natural Resources and Mining", sectors 52-53 "Financial Activities", sector 54-56 "Professional and Business Services", sector 61-62 "Education and Health Services", sector 71-72 "Leisure and Hospitality". : Sector 81 "Other Service Activities" and sector 91 "Public Administration" are excluded.

[8] We took this step because fluctuations in low population cells generate considerable volatility in the estimates.

Using the region-sector weights above, we then create a national net growth series as the weighted average of the industry times region net growth series as follows:

$$G^{t}_{w} = \sum_{sr} W^{srt} S^{srt}_{QBO} G^{srt}$$

We compute region as well as sector-specific series using equivalent weights for industry, and for region. Due to sample sizes, we do not currently include region-sector in the employment index but do include regions (as defined by the Bureau of Economic Analysis) and 20 states.

Benchmarking

For our benchmarking exercise we use the official revenue data for firms with 1 to 9 employees by state, sector, and economy-wide from the 2017 SUSB as the initial value for the index, to which we apply the reweighted firm real revenue growth rates to produce the revenue series of the Small Business Index.

Net growth reweighting

The Small Business Index methodology calculations produce a discrepancy in the total net employment growth between the national series and the aggregated regional or industry series. As described below, we employ a statistical reweighting of the growth rate estimates for the regional and industry series to correct this discrepancy and align with the national data.

Our methodology generates a growth rate series for the US as a whole or by sector, region, or state. From these rates, we produce actual revenue numbers in levels and changes. Discrepancies in the total net growth revenue between the national series and the aggregated industry or regional series can arise. These discrepancies are the result of statistical and rounding errors in the industry and regional growth rate series.[9] To ensure the sub-national series align fully with the national series we recenter the growth rate estimates for the regional and industry series so that the sum of the implied revenue growth from these series equals the national total. More formally, note that the sum of net revenue growth across regions or industries, N_r, need not equal the national totals, N, at time t, or $N_t \neq \sum_r N_{rt}$. To ensure the equality is restored we reweight the industry or region cells using the following weights:

$$W_{rt} = N_{rt} / \sum_{r} N_{rt}$$

[9] Employment is recovered by multiplying the growth rate by the base level employment.

Using these weights, we distribute the national net growth to each region and sector so that the sum of new net revenue growth across regions or industries, $\hat{N}_{rt}$, equals the national net change, or $N_t = \sum_r \hat{N}_{rt}$:

$$\hat{N}_{rt} = N_t * W_{rt}$$

The implied levels for each region and sector in the current period, E_{rt}, are adjusted, $\hat{E}_{rt}$, to reflect the reweighted net growth:

$$\hat{E}_{rt} = E_{rt-1} + \hat{N}_{rt}$$

Finally, the net growth rates for the current period, G_{rt}, are adjusted, $\hat{G}_{rt}$, to reflect the reweighted net growth:

$$\hat{G}_{rt} = \frac{\hat{E}_{rt} - E_{rt-1}}{.5*(\hat{E}_{rt} + E_{rt-1})}$$