Research Examples using QWI Explorer

Using the QWI to Study Startup Firms in the Pittsburgh Metro Area

Suppose we are interested in pursuing a research agenda focused on startup firms in the Pittsburgh Metro Area. We can use QWI data, with tabulations by firm age, to study these new firms.

The QWI provides detailed tabulations by 5 firm age categories: 0-1 years, 2-3 years, 4-5 years, 6-10 years, and 11+ years. For this example, let us define “startup” firms as those that are of age 0-1. With this setup, we can investigate a variety of research questions.

We’ll use QWI Explorer for our analysis: http://qwiexplorer.ces.census.gov/

Research Question 1: Who Works at Startups?

Does the composition of the workforce at startups differ from the workforce at other firms? To answer this, we can compare the distribution of various worker characteristics at startups with the distribution at all firms.

For example, let’s look at the age distribution.

In QWI Explorer, follow these steps:

- Set State to Pennsylvania
- Set Indicator to Emp (Beginning-of-Quarter Employment)
- Change X-Axis to Worker Age
  - Click “Continue” on the conflict box (moving Worker Age from Group to X-Axis)
  - Click the blue text under the X-Axis dropdown, which reads “Eight Age Groups Selected,” and check the box for “All Ages” (so that all boxes are checked)
- Change Group to Firm Age
  - Click “continue” on the conflict box (this will change ownership to Private)
  - Click “close” on the “No Data Available” box (firm age is not available for the most recent quarter. We’ll resolve this by choosing the filter)
  - Click the blue text under the Group dropdown, which reads “Five Firm Ages Selected”, and check the box for “All Firm Ages” (so that all boxes are checked)
- Under Filters:
  - Quarters – click the blue text, deselect 2013Q1, and check all 4 quarters in 2012. This will generate the average employment over these 4 quarters.
  - Sub-State Geography – click the blue text, change the dropdown to Metro/Micro Area, and check the box for Pittsburgh
The screen should look similar to this:

From looking at the table, we can see that the age distribution at startups differs from all firms – the largest Worker Age category for 0-1 Years is 25-34, while the largest category for All Firm Ages is 45-54.

Let’s calculate the percentage of workers that are under age 35, by firm age. Click the Download button, and select “Download Table as Zip.” Open the CSV file.

In a new row, add together employment for ages 14-18, 19-21, 22-24, and 25-34 (should be rows 3-6). Do this for the “All Firm Ages” and “0-1 Years” columns.

In the next row, divide this total by the “All Ages” Row. This gives us the share of workers that are age 14-34 in All Firms and in Startup Firms. The spreadsheet should look similar to this:

<table>
<thead>
<tr>
<th>Age</th>
<th>All Firm Ages</th>
<th>All Firm Ages Flags</th>
<th>0-1 Years</th>
<th>0-1 Years Flags</th>
<th>2-3 Years</th>
<th>2-3 Years Flags</th>
<th>4-5 Years</th>
<th>4-5 Years Flags</th>
<th>6-10 Years</th>
<th>6-10 Years Flags</th>
<th>11+ Years</th>
<th>11+ Years Flags</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Ages</td>
<td>1020781</td>
<td>10</td>
<td>28798</td>
<td>10</td>
<td>28409</td>
<td>10</td>
<td>30001</td>
<td>10</td>
<td>80124</td>
<td>10</td>
<td>853447</td>
<td>10</td>
</tr>
<tr>
<td>14-18</td>
<td>24831</td>
<td>10</td>
<td>1236</td>
<td>10</td>
<td>1081</td>
<td>10</td>
<td>1131</td>
<td>10</td>
<td>2574</td>
<td>10</td>
<td>18806</td>
<td>10</td>
</tr>
<tr>
<td>19-21</td>
<td>51265</td>
<td>10</td>
<td>2219</td>
<td>10</td>
<td>2006</td>
<td>10</td>
<td>2052</td>
<td>10</td>
<td>4726</td>
<td>10</td>
<td>40260</td>
<td>10</td>
</tr>
<tr>
<td>22-24</td>
<td>65788</td>
<td>10</td>
<td>2626</td>
<td>10</td>
<td>2721</td>
<td>10</td>
<td>2538</td>
<td>10</td>
<td>5771</td>
<td>10</td>
<td>52130</td>
<td>10</td>
</tr>
<tr>
<td>25-34</td>
<td>212942</td>
<td>10</td>
<td>6692</td>
<td>10</td>
<td>7614</td>
<td>10</td>
<td>7683</td>
<td>10</td>
<td>18154</td>
<td>10</td>
<td>172799</td>
<td>10</td>
</tr>
<tr>
<td>35-44</td>
<td>196736</td>
<td>10</td>
<td>5266</td>
<td>10</td>
<td>5403</td>
<td>10</td>
<td>5928</td>
<td>10</td>
<td>15334</td>
<td>10</td>
<td>164803</td>
<td>10</td>
</tr>
<tr>
<td>45-54</td>
<td>238594</td>
<td>10</td>
<td>5616</td>
<td>10</td>
<td>5263</td>
<td>10</td>
<td>5923</td>
<td>10</td>
<td>17237</td>
<td>10</td>
<td>204552</td>
<td>10</td>
</tr>
<tr>
<td>55-64</td>
<td>179876</td>
<td>10</td>
<td>4125</td>
<td>10</td>
<td>3211</td>
<td>10</td>
<td>3630</td>
<td>10</td>
<td>12119</td>
<td>10</td>
<td>156791</td>
<td>10</td>
</tr>
<tr>
<td>65-99</td>
<td>50746</td>
<td>10</td>
<td>1016</td>
<td>10</td>
<td>1107</td>
<td>10</td>
<td>1112</td>
<td>10</td>
<td>4205</td>
<td>10</td>
<td>43304</td>
<td>10</td>
</tr>
</tbody>
</table>

This table suggests that startup firms disproportionately employ younger workers: 44% of workers at startups are under age 35, compared to only 35% at all firms.
Research Question 2: Which Industries have the Highest Percentage of Employment in Startups?

Let’s look at the share of startup employment in 3-digit NAICS Subsectors in Pennsylvania.

In QWI Explorer, follow these steps:

- Set State to Pennsylvania
- Set Indicator to Emp (Beginning-of-Quarter Employment)
- Change X-Axis to NAICS 3-digit Subsectors
  - Click the blue text under the Group dropdown, which reads “99 NAICS 3-digit Subsectors Selected, and check the box for “000 All NAICS” (so that all boxes are checked)
- Change Group to Firm Age
  - Click “continue” on the conflict box (this will change ownership to Private)
  - Click “close” on the “No Data Available” box (firm age is not available for the most recent quarter. We’ll resolve this by choosing the filter)
  - Click the blue text under the Group dropdown, which reads “Five Firm Ages Selected”, and check ONLY the boxes for “All Firm Ages” and “0-1 Years” (so only 2 boxes checked)
- Under Filters:
  - Quarters – click the blue text, deselect 2013Q1, and check all 4 quarters in 2012. This will generate the average employment over these 4 quarters.
  - Note that we can not filter by sub-state geography – At the NAICS 3 and NAICS 4 level, firm age and firm size are only available at the state level

The screen should look similar to this:
To find the Subsectors with the highest percentage of startups, we need to move this to a Spreadsheet, calculate percentages, and then sort.

Click the Download button, and select “Download Table as Zip.” Open the CSV file.

Perform the following steps:
- In the next empty column (likely column F, depending on your spreadsheet’s formatting), type “Share 0-1 Years” in the first row.
- In the second row, divide the value in “0-1 Years” by the value in “All Firm Ages” (You should be able to type = D2/B2, although the column letters may differ depending on your spreadsheet).
- After that formula calculates, select that cell, and then click and drag on the bottom right-hand corner of that cell to copy the formula down.
- Finally, select this “Share 0-1 Years” column, click “Sort and Filter,” and then click “Sort largest to smallest”

The results (ignoring the ‘Divide by Zero’ rows) should look like this:

<table>
<thead>
<tr>
<th>All NAICS Subsectors</th>
<th>All Firm Ages</th>
<th>All Firm Ages Flags</th>
<th>0-1 Years</th>
<th>0-1 Years Flags</th>
<th>Share 0-1 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postal Service</td>
<td>13</td>
<td>10</td>
<td>4</td>
<td>10</td>
<td>31%</td>
</tr>
<tr>
<td>Scenic and Sightseeing Transportation</td>
<td>821</td>
<td>10</td>
<td>186</td>
<td>12</td>
<td>23%</td>
</tr>
<tr>
<td>Private Households</td>
<td>4063</td>
<td>10</td>
<td>894</td>
<td>10</td>
<td>22%</td>
</tr>
<tr>
<td>Forestry and Logging</td>
<td>832</td>
<td>10</td>
<td>156</td>
<td>10</td>
<td>19%</td>
</tr>
<tr>
<td>Food Services and Drinking Places</td>
<td>370335</td>
<td>10</td>
<td>29590</td>
<td>10</td>
<td>8%</td>
</tr>
<tr>
<td>Social Assistance</td>
<td>149556</td>
<td>10</td>
<td>11079</td>
<td>10</td>
<td>7%</td>
</tr>
<tr>
<td>Support Activities for Mining</td>
<td>16458</td>
<td>10</td>
<td>1197</td>
<td>10</td>
<td>7%</td>
</tr>
<tr>
<td>Personal and Laundry Services</td>
<td>59391</td>
<td>10</td>
<td>3943</td>
<td>10</td>
<td>7%</td>
</tr>
<tr>
<td>Primary Metal Manufacturing</td>
<td>38650</td>
<td>10</td>
<td>2501</td>
<td>10</td>
<td>6%</td>
</tr>
<tr>
<td>Petroleum and Coal Products Manufacturing</td>
<td>6368</td>
<td>10</td>
<td>376</td>
<td>11</td>
<td>6%</td>
</tr>
<tr>
<td>Accommodation</td>
<td>51417</td>
<td>10</td>
<td>2882</td>
<td>10</td>
<td>6%</td>
</tr>
<tr>
<td>Truck Transportation</td>
<td>61159</td>
<td>10</td>
<td>3236</td>
<td>10</td>
<td>5%</td>
</tr>
<tr>
<td>Support Activities for Agriculture and Forestry</td>
<td>3123</td>
<td>10</td>
<td>164</td>
<td>10</td>
<td>5%</td>
</tr>
<tr>
<td>All NAICS Subsectors</td>
<td>4878959</td>
<td>10</td>
<td>149668</td>
<td>10</td>
<td>3%</td>
</tr>
</tbody>
</table>

Note that we should probably disregard the Postal Service, since its employment levels are so small.

Otherwise, we can note that several industries have particularly high levels of their employment in startups.
Research Question 3: Does Regional Variation in Startup Concentration Predict Employment Growth?

Finally, suppose we want to study the relationship between startup concentration and overall employment growth. For instance, does a large concentration of startups drive employment growth? If so, this would have interesting policy implications – perhaps localities would want to do more to attract and support startups.

To start to answer this question, let’s examine regional variation in the concentration of startup employment. Specifically, we will test whether metro areas with the highest concentration of startup employment in 2002 experienced high employment growth over the next decade.

In QWI Explorer, follow these steps:

- **Set State to Pennsylvania**
- **Set Indicator to Emp (Beginning-of-Quarter Employment)**
- **Change X-Axis to Metro/Micro Areas**
  - Click the blue text under the X-Axis dropdown, which reads “Metro/Micro Areas Selected,” and check the box for “42 Pennsylvania” (so that all boxes are checked)
- **Change Group to Firm Age**
  - Click “continue” on the conflict box (this will change ownership to Private)
  - Click “close” on the “No Data Available” box (firm age is not available for the most recent quarter. We’ll resolve this by choosing the filter)
  - Click the blue text under the Group dropdown, which reads “Five Firm Ages Selected”, and check ONLY the boxes for “All Firm Ages” and “0-1 Years” (so only 2 boxes checked)
- **Under Filters:**
  - **Quarters** – click the blue text, deselect 2013Q1, and check all 4 quarters in 2002. This will generate the average employment over these 4 quarters.

The screen should look like this:
Similar to Example 2, we want to identify the Metro Areas with the highest percentage of startups.

Click the Download button, and select “Download Table as Zip.” Open the CSV file.

Perform the following steps:
- In the next empty column (likely column F, depending on your spreadsheet’s formatting), type “Share 0-1 Years” in the first row.
- In the second row, divide the value in “0-1 Years” by the value in “All Firm Ages” (You should be able to type = D2/B2, although the column letters may differ depending on your spreadsheet).
- After that formula calculates, select that cell, and then click and drag on the bottom right-hand corner of that cell to copy the formula down.
- Finally, select this “Share 0-1 Years” column, click “Sort and Filter,” and then click “Sort largest to smallest”

The resulting output should look like this:

<table>
<thead>
<tr>
<th>All Firm Ages</th>
<th>All Firm Ages Flags</th>
<th>0-1 Years</th>
<th>0-1 Years Flags</th>
<th>Share 0-1 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York-Northern New Jersey-Long Island, NY-NJ-PA (PA part)</td>
<td>6768</td>
<td>10</td>
<td>525</td>
<td>10</td>
</tr>
<tr>
<td>Harrisburg-Carlisle, PA</td>
<td>252000</td>
<td>10</td>
<td>13761</td>
<td>10</td>
</tr>
<tr>
<td>Sunbury, PA</td>
<td>24818</td>
<td>10</td>
<td>1319</td>
<td>12</td>
</tr>
<tr>
<td>Lewistown, PA</td>
<td>12992</td>
<td>10</td>
<td>644</td>
<td>12</td>
</tr>
<tr>
<td>Indiana, PA</td>
<td>24312</td>
<td>10</td>
<td>1196</td>
<td>10</td>
</tr>
<tr>
<td>Oil City, PA</td>
<td>17847</td>
<td>10</td>
<td>876</td>
<td>10</td>
</tr>
<tr>
<td>East Stroudsburg, PA</td>
<td>39113</td>
<td>10</td>
<td>1884</td>
<td>10</td>
</tr>
<tr>
<td>Selinsgrove, PA</td>
<td>14556</td>
<td>10</td>
<td>699</td>
<td>12</td>
</tr>
<tr>
<td>Johnstown, PA</td>
<td>47894</td>
<td>10</td>
<td>2291</td>
<td>10</td>
</tr>
<tr>
<td>Warren, PA</td>
<td>12758</td>
<td>10</td>
<td>574</td>
<td>10</td>
</tr>
</tbody>
</table>

Now, let’s focus on the 5 Metro areas with the highest percentage of startup employment. Let’s examine how much their employment grew between 2002 and 2012. Here are the steps in QWI Explorer:

In QWI Explorer, follow these steps:
- **Set State to Pennsylvania**
- **Set Indicator to Emp (Beginning-of-Quarter Employment)**
- **Change X-Axis to Year/Quarter**
  - Click the blue text under the X-Axis dropdown, which reads “Eight Year/Quarters Selected,” and check ONLY the boxes for 2002Q1 and 2012Q1.
- **Change Group to No Group**
- **Under Filters:**
  - **Firm Ownership** – click the blue text, and select the radio button for “All Private Ownership.” We want to focus our analysis on private firms only (Since we identified startup concentration by region for only private firms, we want to examine employment growth for private firms – important to be consistent)
  - **Sub-State Geography** – click the blue text, change the dropdown menu to micro/metro area, and select the 5 metro areas that had the highest concentration of startup employment in 2002:
    - Harrisburg-Carlisle
    - Indiana
    - Lewistown
    - New York-Northern New Jersey-Long Island
    - Sunbury

This will generate the total employment for these 5 metro areas.
Employment in these metro areas grew from 310,742 in 2002Q1 to 331,786 in 2012Q1.

To calculate the percentage growth, we can download this to a spreadsheet and divide the two rows. You would find that \( \frac{331,786 - 310,742}{310,742} = 6.8\% \), suggesting that total employment grew by 6.8% in these regions.

How does this compare to the state of Pennsylvania as a whole? Simply remove the filters: click on the blue text next to “Sum of Metro/Micro Areas,” and click “Check None.”

The screen should now look like this:

Notice that employment in the state of Pennsylvania grew from 4,703,170 in 2002Q1 to 4,775,714 in 2012Q1. Calculating the percentage growth gives us \( \frac{4,775,714 - 4,703,170}{4,703,170} = 1.5\% \). This suggests that employment in the 5 metro areas with the highest concentration of startup firms in 2002 grew over four times faster than employment in the entire state (6.8% vs 1.5%)!!

Remember, correlation is not causation – we can’t say that the concentration of startups necessarily caused the higher rate of employment growth, only that a correlation exists. Still, this is an interesting result that can motivate future research.