Outside of the Box Use of Administrative and Wage Data in Texas

Adam Leonard, Chief Analytics Officer, Texas Workforce Commission

Introduction: This workshop is designed to consider new empirical approaches to the grouping of firms into industries. One possible approach is to trace investments in critical technologies by tracing the flows of research funded individuals to the private sector. Given its explosion into the collective consciousness and the transformative effect it will have on the economy, Artificial Intelligence provides a highly relevant use case for examination.

One potential way is to make use of the universe wage records filed each quarter by employers to state unemployment insurance agencies to help them administer their programs. These records are a potential treasure trove of data that can track the transmission of ideas and their spread through the economy.

Starting with federal grant expenditure data (that is, administrative data), it is possible to identify who is funded to do research from a grant. If we were to connect the data identifying those doing research on the grant to the university student administrative data or to a State Longitudinal Data System (SLDS, sometimes referred to as a P20W system). This allows us to track the spread of ideas from faculty to students.

It is then possible to move longitudinally to trace where these grant workers and students go when they leave the university and become employed (or start a business and employ others) by looking at wage records over time. Further, by starting back with the grants, we're able to understand how these technologies are spreading through industries through their hiring in ways that are not always possible through surveys conducted in accordance with U.S. Bureau of Labor Statistics (BLS) standards such as through the Quarterly Census of Employment and Wages (QCEW). Certainly, the current BLS NAICS code classification for AI is implausibly low in Texas – only 298 firms are so identified with a total of 1021 workers.

What is important is though QCEW includes detailed industry information and "primary product or service" for a firm, that data is limited in telling the story of how a new technology is being developed or utilized by those firms. Further, the primary product or service data is woefully under-reported by employers making it, at best, slightly useful.

Most of the biggest players in the AI space are established companies that are adding AI to their portfolio – their QCEW profile does not reflect this shift. Other companies that are investing heavily in AI are not even in the traditional tech sector but rather are in finance, insurance, bioscience and pharmaceuticals, and oil and gas. Through wage records, we can measure more directly the type of investment firms are making in AI – by looking at the number of employees they are hiring with AI backgrounds and how that relates to the number of employees overall and how those numbers change over time.

In this white paper I will discuss the feasibility of such an empirical approach using Texas workforce data.

The Texas Experience: The Texas Workforce Commission has long realized that its greatest untapped resource was its data. Data represents our past, informs our present, and can shape our future. It has investing heavily in trying to tap into that resource to help Texas employers, individuals, families, and

communities achieve and sustain economic prosperity; that is, Data for Prosperity (TM pending). The remainder of this piece will describe ways that the state has used administrative and wage data for alternative purposes in the past and how we created a very limited test of some of the ideas being considered in this workshop for tracking the spread of influence of ideas to particular firms and clustering them into "industries of ideas."

As part of that work, TWC has been using wage records in ways to better understand our program results, the experiences of employers and others we served, and the labor market more broadly. This moves beyond simply measuring performance or being limited to BLS standards in our work. TWC has long used wage data to better understand employment patterns in industries. One key concept that is currently being built out is focusing on new employment connections between workers and employers (presumed to take place when we find a worker reported for an employer in a quarter and that worker was not reported by that employer in the prior two quarters).

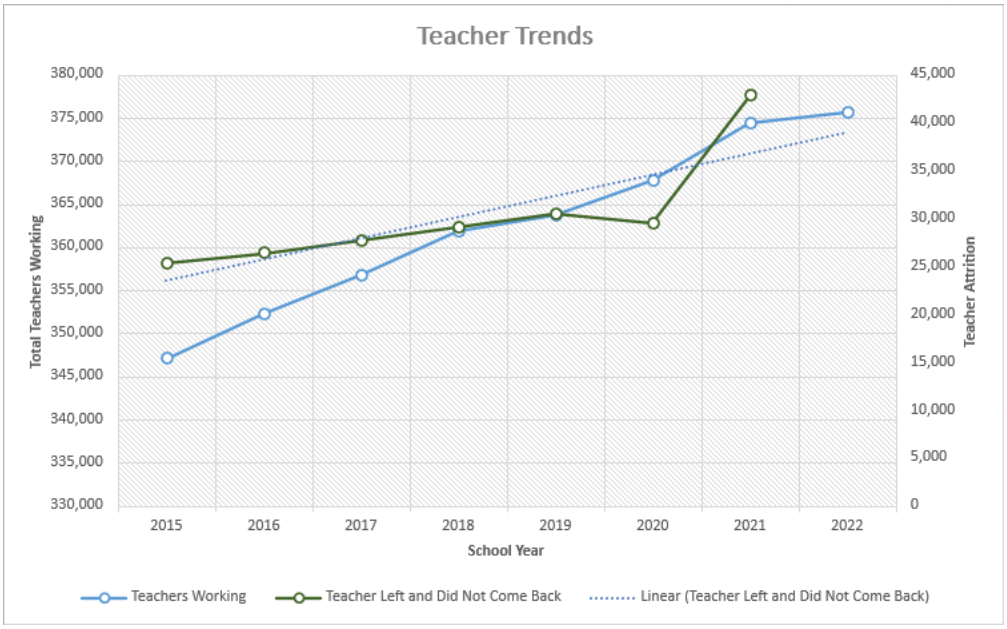
We then look to see what percentage of those new employment connections are maintained for at least two subsequent quarters. We're able to examine this data by industry and in doing so can better characterize the nature of the industry and identify positive and negative outliers. It also creates an objective, independent picture of natural churn in the full labor market and gives programs a way to evaluate their success in helping people retain employment.

Beyond that, TWC has found that combining administrative data and wage data can be extremely helpful in evaluating narratives. For example, in the last year, the media had numerous stories about difficulties schools reported finding teachers and the large number of teachers leaving the profession.

- Struggling to find teachers close to home, some Texas schools are looking oversees for help, Texas Tribune
- Extended Coverage: Texas Teachers are leaving the profession in record numbers. They told us why, CBS News
- Texas tried to fix its teacher shortage by lowering requirements – the result was more new teachers, but at lower salaries, The Conversation

TWC, using teacher data from the Texas Education Agency (TEA) through the Texas Tri-Agency Workforce Initiative, was able to identify teachers who left the profession each year from 2015 through 2022 and then see where they went through wage record analysis. This analysis contextualized and tempered the prevailing narrative in several ways.

First, the data showed that the number of Teachers working in Texas grew every year over the period examined and that the growth curve was largely unchanged from pre to post pandemic(indeed there was a significant bump in 2022).



Second the data showed that a high percentage of teachers "leaving" were still employed in Education Services. This caused us to go back to TEA for more data on the roles these former teachers still employed in Education Services were holding. The vast majority of them were in roles that were more senior in nature such as moving into a Assistant Principal (or full Principal), Counselor, Department Head, or Teacher Supervisor/Facilitator.

Third, a significant portion of the teachers who left the profession did not show up in wage records at all, suggesting that some portion of them were retiring – matching data to the Texas Retirement System might provide additional clues in that regard.

Together, this data does not refute the difficulty that many school districts may have with finding enough teachers. However, it does provide a more nuanced picture and extent of the issue as well as the factors contributing to it.

What makes this work interesting is that when TWC recently wanted to examine its own turnover problem – in our case with vocational rehabilitation counselors – we set up an examination similar to the teacher work. While the work is not complete, it demonstrates the transmission of ideas through connection in the workplace. The people working on the VR counselor problem were discussing their issue with the people who did the teacher analysis and the technique was passed on to the new team organically – it wasn't the result of the division director instructing them how to do it – it was the result of peer to peer communication and collaboration.

TWC also has begun experimenting more directly the industry of ideas approach using a very quick and simple proof of concept. Given the short timeframe, we started with employees at the University of Texas at Austin who were employed in CY2021 and who were not employed at UT in CY2022. We then isolated the employees working in CY2022 with quarterly earnings of $50,000 or more (with the assumption that people previously working at a University and then immediately earning $200,000 a year or more are very likely to be in high demand fields – such as AI or computer science – this is imperfect but workable for the exercise). There were 140 such individuals.

We then examined the employers they became employed with and the industries those employers were in. While we can't be sure all 140 former employees were in AI or other advanced computer work, Many were at firms that identified as being in the "Custom Computer Programing" industry. More interesting was the other industries we found these workers in. Some, such as "Offices or Lawyers," were probably not relevant but many more likely were. For example, many were working for some of the biggest e-commerce, internet, AI hardware, biochemical/pharmaceutical, and materials science firms in the world – all of which makes sense given the potential for AI to transform those industries.

Once those firms are identified, it is then possible to examine the educational background of all the workers in those firms. States have long used State Longitudinal Data Systems (SLDS, also known as P20W system) to track students from K-12, through post-secondary education and then through their careers. These systems combine administrative data (from schools) about students, where they went to school, what they studies, and what credentials they achieve with wage data to show where they are working, how much their earnings are, and how those earnings change over time. But that has been primarily as a education evaluation system and a way to help students make informed decisions about which schools to attend and programs to enroll in.

The combination of administrative and wage records provide an avenue for rich examination of employment and earnings trends and the spread of ideas and impact of federal investments through the migration of people through their careers and firms through their hiring and employment practices. If we were to combine SLDS with grant tracking data we will be able to use them to do a more macroeconomic analysis of industries and technologies and how they change over time.