NBER WORKING PAPER SERIES

ARE PINK SLIPS BETTER THAN FLU SHOTS? THE EFFECTS OF EMPLOYMENT ON INFLUENZA RATES

Sara Markowitz
Erik Nesson
Joshua Robinson

Working Paper 15796
http://www.nber.org/papers/w15796

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2010

We would like to thank Barry Hirsch, David Frisvold, and participants at Emory University seminars for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Are Pink Slips Better Than Flu Shots? The Effects of Employment on Influenza Rates
Sara Markowitz, Erik Nesson, and Joshua Robinson
NBER Working Paper No. 15796
March 2010
JEL No. I0,J0

ABSTRACT

The seasonal influenza virus afflicts between five and twenty percent of the U.S. population each year, imposing significant costs on those who fall ill, their families, employers, and the health care system. The flu is transmitted via droplet spread or close contact, and certain environments, such as schools or offices, promote transmission. In this paper, we examine whether increases in labor market activities are associated with an increased incidence of the flu. Flu data come from the Centers for Disease Control. We check the robustness of our results using unique data from Google Flu Trends. Using a first-difference two stage least squares estimation approach, we find that a one percentage point increase in the employment rate increases the number of influenza related doctor visits by about 8.1 additional flu-related doctor visits per 1000 doctor visits for all causes. To put this in perspective, on average, 33 additional people out of every 100,000 new employees will have a flu-related doctor visit. The results are robust across several specifications.

Sara Markowitz
Department of Economics
Emory University
1602 Fishburne Dr.
Atlanta, GA 30322
and NBER
sara.markowitz@emory.edu

Joshua Robinson
Department of Economics
Emory University
1602 Fishburne Dr.
Atlanta, GA 30322
jjrobi2@emory.edu

Erik Nesson
Department of Economics
Emory University
1602 Fishburne Dr.
Atlanta, GA 30322
enesson@emory.edu

1 Introduction

The seasonal influenza epidemic imparts surprisingly large costs on the economy. According to the Centers for Disease Control and Prevention (CDC), the flu infects between five and twenty percent of the population each year, causing mild to severe illness. Complications from the flu hospitalize more than 200,000 Americans annually, and on average 36,000 people die from these complications every year. Molinari et al. (2007) estimate the economic burden of the flu at nearly 90 billion dollars per year.

The flu is transmitted from sick to well individuals when a healthy person either touches a surface containing the flu virus or inhales the virus after an infected person coughs or sneezes. As we discuss in detail below, many of the characteristics or environments associated with modern daily work life may encourage the transmission of the flu. These characteristics include commuting via public transportation, working in climate controlled indoor offices, sharing workspaces, and placing young children in daycare. If the features of the working environment promote the spread of the flu, then being out of the labor force and away from these environments could help reduce the spread of the flu. In this paper, we explore this proposition and measure the degree to which the transmission of the flu is related to labor market activities.

In our main results, we use data from the CDC in conjunction with two measures of labor market activity. We estimate a dynamic panel data model using the Anderson and Hsiao (1981) FD2SLS estimator for the 1997-1998 to 2007-2008 flu seasons. We find evidence that increased labor market activity indeed aids the spread of the flu. We find that a one percentage point increase in labor market activity is associated with about a 0.81 percentage point increase in the incidence of the flu, or 33 additional people out of every 100,000 new employees. We confirm these results with data from Google Flu Trends, a state-level data set that proxies flu prevalence by the prominence of flu-related web searches.

1.1 Background

According to the CDC, there are three types of influenza viruses: A, B, and C. Types A and B are responsible for the seasonal epidemics in humans. Influenza A can be broken down into subtypes based on the proteins present on the surface of the virus. There are two such subtypes currently found in humans, each of which has numerous unique strains: A(H3N2) and A(H1N1), a novel strain of which is responsible for the current "swine" flu pandemic. Influenza B does not have subtypes, but it still has many different strains. The influenza virus mutates at a very rapid rate, resulting in a high number of unique strains of the virus. A large number of unique strains means that people cannot become indefinitely immune to the virus from adaptive immunity or through vaccination (CDC, 2010). Having recovered from a past flu infection or having been vaccinated against known strains of the flu does not protect against infection from new strains of the virus. When new flu strains emerge, it can lead to flu pandemics, during which widespread infection, hospitalization, and death are common.

Influenza epidemics are seasonal, occurring mainly in the winter months. The CDC actively monitors flu activity during the flu season, defined by the CDC as between October and May (CDC, 2010). There is no consensus among the scientific community as to why flu outbreaks occur mainly in the winter months (Cox and Subbarao, 2000). Since the 1982-1983 flu season, the peak month of flu activity occurred in February nearly half of the time. However, the peak month of activity can occur as early as November, and regularly occurs in March.

The mechanism for flu transmission is sick-to-well transmission. That is, healthy people become infected with influenza by coming into contact with infected individuals. According to the CDC, the transfer of the virus occurs through droplet spread. The influenza virus is present in the saliva and mucus of infected individuals, and when an infected person coughs

or sneezes, the virus is introduced into the air in aerosol form. The virus can either be inhaled when airborne, called aerosol transmission, or introduced into the body after touching a surface containing the droplets, called close-contact transmission. Experimental evidence suggests that, under laboratory conditions, aerosol transmission is the most successful mechanism of infection (Tellier, 2006). However, the virus can only survive for 16 hours in aerosol form. On the other hand, the virus can survive for up to 48 hours on a non-porous, dry surface. This evidence—coupled with experimental evidence showing up to 60 percent of the objects surrounding an infected person have an infectious concentration of the virus—leads some researchers to believe close contact transmission is primary mechanism of flu spread (Brankston et al., 2007; Tellier, 2006).

Both school and places of employment offer the close contact between individuals and commonly touched surfaces conducive to the spread of influenza. Cox and Subbarao (2000) report that the highest attack rates of influenza occur among school age children, and that school absenteeism is typically followed by work absenteeism. Individuals in families with school-age children (or children in daycare) are more than twice as likely to become infected with influenza, which suggests schools are a primary environment for the spread of the flu. But because schooling is compulsory, the large variability in flu incidence between regions and both within and across flu seasons is unlikely to be explained by school enrollment. Employment provides an analogous environment for adults, and employment rates vary significantly across regions and time.

A person who is employed could be much more likely to come into close contact with the influenza virus than individuals out of work or out of the labor force. Since the employed typically commute to and from the workplace, they are more likely to use public transportation or car pools, which are favorable environments for flu transmission. According to the 2000 US Census, over six million Americans commute using public transportation, and another fifteen million commute by car pool. This puts roughly 17 percent of working Americans at

risk of catching the flu in these manners. Moreover, the majority of Americans work indoors. Working environments like offices are ideal for both aerosol and close-contact transmission of the influenza virus. In addition to the close proximity of workers, Tellier (2006) reports that the influenza virus in aerosol form survives better in low humidity environments. Thus, a climate-controlled office building would be a conducive environment for aerosol transmission. Commonly touched surfaces such as keyboards, water coolers, and communal restrooms also promote close-contact flu transmission. Finally, employed individuals are more likely to put young children in daycare. As noted above, school and daycare are fertile environments for the spread of the flu.

It is worth noting that there are alternative (and not generally-accepted) views concerning influenza transmission. This literature suggests that influenza does not behave like a typical infectious disease, and sick-to-well transmission may not adequately explain the epidemiology of the flu. One theory suggests epidemics of the flu are due to vitamin D deficiency and its importance to innate immunity (Cannell et al., 2008).[1] Although this theory of flu transmission is very different from the generally accepted view, influenza transmission could still be highly influenced by the economic environment. Since the majority of Americans work during the daylight hours, which limits sun exposure, it follows that vitamin D deficiency may increase in prevalence during times of increased indoor employment. Additionally, person-to-person contact is still necessary for the flu spread, even if it is not responsible for the outbreak of an epidemic. Thus, in the event that this theory of flu transmission has more substance than is currently attributed to it, we should still see an effect of employment on flu rates.

This paper adds to the literature concerning the effect of macroeconomic conditions on health. The bulk of this literature relates local unemployment rates to a variety of health outcomes. For example, Ruhm (2000) finds that higher state-level unemployment rates are

[1]See White (2008) for a description of vitamin D's importance for innate immunity.

associated with decreased mortality rates, including deaths from influenza and pneumonia. Ruhm and Black (2002) find that people consume less alcohol in bad economic times, and Ruhm (2003) finds a counter-cyclical variation in physical health measures, which is particularly pronounced among working-age men. The same relationship with macroeconomic conditions is confirmed for smoking and excessive weight (Ruhm, 2005), coronary heart disease (Ruhm, 2007), and for mortality rates for all OECD countries (Gerdtham and Ruhm, 2006). There is some evidence to suggest that these analyses may be sensitive to the business cycle measure used, and that the probability of death may in fact go up during a recession (Gerdtham and Johannesson, 1999). Our research question here is not concerned with unemployment or recessions per se, rather we are interested in the effects that aggregate employment rates or job-seeking may have on the spread of a common but potentially deadly virus.

2 Methodology

We exploit regional variations across time in both influenza incidence and labor market activity to test the theory that increases in employment levels increase individuals' risk of exposure to the influenza virus. We examine the spread of the flu during the flu season with a dynamic panel data model. We only focus on the flu season because, first, there is virtually no flu activity during the summer, and second, the CDC does not monitor flu activity during the summer for most of our time period. For region j, flu season s, and month m, we estimate the following panel data model:

$$\log \mathrm{F}_{jsm} = \alpha + \gamma E_{jsm} + \delta \log F_{jsm-1} + \beta X_{jsm} + R_j + \tau_t + \epsilon_{jsm} \tag{1}$$

In Equation 1, the dependent variable is the natural log of a measure of flu prevalence (F_{jsm}). Preliminary regressions using a linear dependent variable showed heteroskedastic errors increasing as the flu rate increased. To address this problem, we use a log-linear specification with population weights.[2] E_{jsm} is a measure of labor market conditions (details are discussed below). Since influenza is a communicable disease, the number of individuals that will be infected depends on the number that were previously infected. We include a lagged dependent variable, $\log F_{jsm-1}$, to account for the number of infectious people in the region during the previous month. X_{jsm} is a vector of regional demographic characteristics used as proxies for the susceptible population and other measurable factors affecting the spread and reporting of the flu, and R and τ are region and time fixed-effects. We include region fixed-effects to account for time-invariant unobservable characteristics that may affect the flu spread within a region. Additionally, the severity of the flu depends heavily on the particular strains that are active, whether those strains are successfully targeted by the flu vaccine, and how many people get vaccinated. In order to account for this, we include flu season fixed-effects, which are indicators for each October to May timespan. As discussed further below, we also include month trends and month indicator variables in certain specifications.

One problem with estimating Equation 1 is that the coefficient on a lagged dependent variable is biased and inconsistent in a fixed-effects panel model. As Nickell (1981) points out, the lagged dependent variable is correlated with the mean of the error term. This problem becomes more severe in instances where the number of cross-sectional observations is small (Beggs and Nerlove, 1988). In order to deal with this problem, we employ the Anderson and Hsiao (1981) first difference two stage least squares estimator. To construct the FD2SLS estimator, Equation 1 is differenced, removing any variables that do not vary

[2]Running our models using linear dependent variables does not change our results.

within a flu season.[3] However, first differencing creates a new endogeneity problem in that the first difference of the lagged flu incidence ($\log F_{jsm-1} - \log F_{jsm-2}$) is correlated with the first difference of the error term ($\epsilon_{jsm} - \epsilon_{jsm-1}$). To deal with this problem, the second lag of the flu incidence ($\log F_{jsm-2}$) is used as an instrument for the lagged first difference.[4] Since the first differencing removes region-level flu season effects, this estimation strategy also eliminates any need to control for unobservables that vary by flu season. The FD2SLS model we estimate is

$$\Delta \log \mathrm{F}_{jsm} = \gamma \Delta E_{jsm} + \delta \Delta \log F_{jsm-1} + \phi \Delta X_{jsm} + \Delta \epsilon_{jsm}, \tag{2}$$

where all the variables are as in Equation 1, except that X_{jsm} includes only those variables that vary more than once a flu season.

3 Data

We use two monthly variables constructed from the Bureau of Labor Statistics to measure the conditions in the labor market. Each month, Census workers interview people in 60,000 households about their labor market activities during the week that includes the 12th day of the month. These values are adjusted for sample demographics and extrapolated to state-level or region-level employment figures.[5] The first variable we use to measure labor market activity is a constructed employment rate of the percent of the civilian population currently

[3]Because our model runs over each flu season and not as a continuous panel, first differencing removes any flu season-specific variables in addition to any time-invariant, region-specific fixed-effects.

[4]The second lag should be highly correlated with the lagged first difference, but is also exogenous to the first difference of the errors($\epsilon_{jsm} - \epsilon_{jsm-1}$) since future shocks in flu incidence cannot have any impact on past flu incidence.

[5]For more information on the construction of the BLS data, see http://www.bls.gov/cps/cps_htgm.htm.

working. The employment rate is

$$EMP = \left(\frac{\text{employed population}}{\text{non-institutional civilian population}} \right) \times 100.$$

We use this employment rate rather than the well-known unemployment rate because the unemployment rate denominator is the labor force and not the working-age population. Since the unemployment rate denominator fluctuates as people enter or exit the labor force, the unemployment rate will not fully capture the changes in the percentage of the population working.

One potential problem with the employment variable is that people not working but in the labor force could be engaging in many job-searching activities that could spread the flu. As an alternative measure, we calculate the percent of the civilian population in the labor force,

$$PLF = \left(\frac{\text{population in labor force}}{\text{non-institutional civilian population}} \right) \times 100.$$

This labor force participation rate should pick up the effects of people engaging in some sort of labor-market activities on the spread of the flu. However, since some individuals may only search for jobs by web searching, sending out résumés, or other forms of search that do not involve person-to-person contact, we expect the point estimates for the labor force participation variable to be smaller in magnitude than that for the employment rate.

Our primary measure of flu prevalence comes from the CDC. The CDC heads an influenza surveillance system comprised of, among other surveillance measures, an Outpatient Influenza-like Illness Network (ILINet). The ILINet consists of some 2,400 health care providers across all fifty states and Washington D.C. Each week during the flu season, which runs from October through May, providers voluntarily submit the number of patient visits for all causes and the number of patients with an influenza-like illness (ILI). The symptoms

necessary for an ILI diagnosis include a fever greater than 100 degrees Fahrenheit, a cough and/or sore throat, and a lack of a known cause besides influenza. The CDC aggregates the data to the Census Division level and provides weekly data for each flu season from the 1997-1998 season through the current season.[6] Table 1 shows summary statistics for the flu variable and all other variables. On average, during the flu season 2.21 percent of doctor visits are flu-related.

There are some potential pitfalls with this flu measure. First, the location and coverage of the health providers is unknown. There may be a bias towards urban locations and larger, well-staffed medical providers. Second, the reporting may not accurately reflect the severity of the flu. It may be that people visiting the doctor with flu-like symptoms are more from populations at risk from serious complications arising from the flu, such as children, the elderly, and people with concurrent illnesses. Since these populations are less likely to work, it may bias our estimates towards zero. Doctors' reactions may also change based on the severity of the flu. It may be that doctors diagnose and report patients with the flu more often during the height of the flu season. As long as the measurement error from reporting is not systematically related to employment, our estimates will be unbiased, but the standard errors will rise.

Since the flu data are reported on a weekly basis and our other variables are reported on a monthly, quarterly or yearly basis, we aggregate the flu variable to a monthly level. We generate the monthly flu rate to match the time period between the measurement of our labor market variables and the measurement of the flu variable. According to the CDC, an infected person may not show symptoms for up to four days after infectious contact and is capable of infecting others for up to a week after symptoms start. There is also some lag between when symptoms start and when that person visits a physician regarding his or her

[6]Unfortunately, we were not able to obtain the state-level data from the CDC. Also, we do not use the CDC data after the 2007-2008 flu season because the definition of the Census Divisions change.

illness. Given these timing issues, it seems likely that a week or more would pass before the effect of a change in employment could be seen in a change in flu prevalence. In our main results, we use the average of the weekly CDC flu prevalence measures for the weeks starting the fourth week of each month and ending the fourth week of the next month. Since the employment variables are measured over the week containing the 12th day of the month, the measurement of the flu starts, on average, a bit more than a week later. The robustness of our results to different advancement periods and aggregation measures is presented in Section 5. We also present results using a state-level measure of flu prevalence based off of web searches from Google Flu Trends in Section 5.3.

We include in all models the natural log of per-capita income for each census division reported quarterly from the Bureau of Economic Analysis. This clarifies the interpretation of our parameter of interest. We want to capture how changing employment conditions affect the spread of influenza, not how changes in income affect the spread of influenza.

Other variables of interest are only available annually.[7] While this lack of variation prevents us from including these variables in the FD2SLS models, we do show them in OLS models in Section 5.4. These variables include two annual education measures: the percents of the population with at least a high school degree and at least a bachelor's degree. To account for how likely people are to go to a doctor with flu-related symptoms, we include the annual percent of the population with health insurance, the annual percent of the population that is over sixty-five, and annual the percent of the population that is under fifteen. To account for reasons people may be in close contact other than employment, we include an annual measure of the percent of the population that lives in a rural area. One last annual variable is the percent of the population that receives a flu shot. To construct our flu shot variable, we use the Behavioral Risk Factor Surveillance System (BRFSS) telephone

[7]To make the annual variables analogous to a flu season, we construct a weighted average of the two annual variables covering each flu season using the number of months as the weight.

questionnaires. The BRFSS contains a question about whether the individual has receive a flu shot in the previous twelve months and the date of the interview. We assume that individuals do not receive flu shots earlier than October, which allows us to construct a percentage of respondents who received a flu shot for each region in each flu season. Although monthly data would be ideal, we could only construct region-level data for each flu season.

4 Results

Table 2 displays results from our main regression models using the two different measures of labor market activity. All models use the Anderson-Hsiao FD2SLS methodology to correct for the endogeneity arising from the lagged dependent variable. Since our dependent variable is a rate, we weight our regressions by the Census Division population. The standard errors in all models presented below are adjusted for intra-regional correlation according to Huber (1967). The first four models in the table use the employment rate as the measure of labor market activity, and the last four models use the percent of the population in the labor force as the measure of labor market activity. Aside from the measure of labor market activity, the base models include the lagged dependent variable and the log of the per-capita income as independent variables. Since the FD2SLS estimator excludes all variables that do not vary more than once a flu season, we necessarily exclude other explanatory variables which vary annually.

One concern is that the occurrence of influenza follows a fairly regular trend each season. The number of reported flu cases rises steadily, usually peaking in February, before declining each month through May. If there are similar trends in employment, then our results may be spurious. To account for this, we model the flu epidemic in three ways. We first use linear and quadratic month trends. Second, we generate two indicator variables for the time periods before February and after February, leaving the month of February as the base

category. Lastly, we include month fixed-effects to test for any unique monthly trends in the flu epidemic.

All models show a positive effect of employment activity on flu incidence, and nearly all the coefficients are statistically significant at the one percent level. There is some reduction in the magnitude of the coefficients when month effects are included, and the employment rate becomes insignificant. We attribute the lost significance in the month fixed-effects model to a lack of remaining variation in our employment variables. In our preferred model, including the two indicator variables for before and after February, we find that a one percentage point increase in the employment rate increases the incidence of the flu by 36 percent. While an effect of 36 percent seems large, the average ILI measure is just over 2 percent. Thus, a one percentage point increase in the employment rate translates to 8.1 additional flu-related doctor visits per 1000 doctor visits for all causes. Using the the labor force participation rate, we find that a one percentage point increase in labor force participation increases the incidence of the flu by 44 percent. This translates to about 9.7 additional flu-related doctor visits per 1000 doctor visits for all causes. Note that a one percentage point change in employment or labor force participation over the course of a month is quite large. The average change in the employment rate from one month to the next in our sample is 0.3 percentage points, and the average change in labor force participation is just under 0.2 percentage points.

Another way to interpret these results is that for every additional 100,000 individuals employed, we would observe 3.3 new flu cases per 1000 doctor visits. On average, approximately 10,000 out of every 100,000 people visit a general practitioner every month (Cherry et al., 2006). Thus, of the newly employed, we expect 33 additional flu-related doctor visits above the mean, or a new total of 2,246 cases for the 100,000 individuals.[8]

[8]An increase in employment of 100,000 people is slightly above the average monthly net change in employment of 73,000 people per census division.

5 Robustness Checks

5.1 Lagged Effect of Employment Conditions on Influenza

Recall from Section 3 that our flu prevalence measure is constructed by taking the average of the flu prevalence measure starting in the fourth week of each month. In Table 3, we test the sensitivity of our estimates to changing the starting week of the dependent variable. To test the sensitivity of measuring the flu variable starting in the fourth week of each month, we estimate the FD2SLS models with flu-week advances ranging from zero to six weeks. The row labelled "Starting in Week 4" repeats the coefficients from Table 2.

If our assumptions about flu spread and reporting are correct, we expect to see an increase in the size of the coefficients followed by a decline over the course of the advancement periods. In Table 3 the coefficients are small (and less precise) when no time gap is in place. The coefficients for both the employment rate and labor force participation rate increase when the gap increases, peaking at four weeks in most specifications. The coefficients in the employment specifications decrease for the five and six week advances, while the coefficients in the percent labor force specifications do not fall. The lingering effects of a shock to flu prevalence points to the contagious nature of the flu.

5.2 Alternative Construction of the Dependent Variable

One potential problem with our dependent variable is that labor market conditions may affect the rate at which people go to the doctor.[9] If people go to the doctor less when they are unemployed, the number of flu-related physician visits could be likewise affected. To check, we ran a regression of the log of the number of physician visits on the two employment

[9]Another potential problem is that the spread of influenza may affect employment status. For example, someone may be fired for being sick too often, or someone may leave their job to take care of a sick parent, spouse or child. In the aggregate, these sources of bias are likely to be small.

variables and other controls. We found that as employment activity increases, the number of physician visits increases by a very small, but statistically significant amount. Our base models account for this potential bias by using as the denominator for the dependent variable the number of physician visits rather than the population. Thus, both the numerator (number of flu-related physician visits) and the denominator (total number of physician visits) are affected by a change in insurance prevalence in the population. We investigated the potential bias by regressing the log of the number of flu-related physician visits on the variables in our preferred model, but adding in the log of the total number of physician visits as an explanatory variable. The coefficients on the employment measures remain positive and statistically significant.

5.3 Data from Google Trends

Another concern is that the CDC data are only available at the Census Division level. To test the significance of this aggregation, we employ a state-level data set generated by Google Flu Trends available from the 2003-2004 flu season to the 2008-2009 flu season. Google uses an algorithm to choose the combination of search queries that best predict the CDC flu severity variable. Google then uses the parameters from this prediction to create a state-level flu severity measure. When evaluated at the Census Division level, the Google data tracks the CDC data very well, but due to reporting lags in the CDC data the Google data predicts trends about one to two weeks earlier than the CDC data. Thus, we use the log of the average predicted percent of flu-related physician visits beginning the second week of each month as our measure of flu severity, two weeks earlier than the CDC advancement period. Note that because the Google data is based on web queries, it acts as another check on the endogeneity problem discussed previously. For detailed information on the construction of the Google Flu Trends data, see (Ginsberg et al., 2009).

Regressions using the Google data are estimated using FD2SLS and the same control variables as our regressions that use the CDC data. The difference is that the Google data is at the state level rather than the regional level and over the 2003-2004 to 2008-2009 flu seasons instead of the 1997-1998 to 2007-2008 flu seasons. The fixed-effects and other control variables are also measured at the state level. Results from the Google data are displayed in Table 4. For the base regressions and the regressions including either quadratic month trends or indicator variables for before and after February, the sign and magnitude of the coefficients of interest stay positive, statistically significant and of comparable magnitude to the regressions using the CDC data. For the models including month fixed effects, the coefficients lose precision. As with the CDC data, we believe this is due to the month effects explaining nearly all of the variation in this data.

5.4 Other Demographic Variables

In the FD2SLS models, we exclude certain demographic characteristics of interest because the model eliminates any variables that do not change more than once a flu season. In this section we re-estimate the base models from Tables 2 and 4 including these other demographic variables to see if their inclusion has any effect on the significance of our results. To estimate these parameters, we use weighted OLS, and Table 5 shows the results of these regressions. All models include state/region fixed-effects, and state/region-specific linear flu season trends, necessarily ignoring the endogeneity of the lagged dependent variable.

The coefficients on the employment variables remain positive and statistically significant at the five percent level for three of the four specifications. The strong positive relationship of previous flu incidence and current flu rates remains in WOLS models as well. The signs of many of the coefficients, however, are quite unexpected. It is somewhat surprising that the coefficient on the percentage of the of the population without health insurance is negative

in all four specifications (although only significant at the 10 percent level using the CDC data and insignificant when using the Google data). This may be because populations more prone to illness may also be more likely to purchase insurance. Also surprising is the percent of the population receiving the flu shot has no negative (significant) effect on flu rates. We caution that the construction of this variable may not be timed well with the flu data. It is also possible that the coefficients may reflect reverse causality, where people react to high incidence of influenza by getting a flu shot. Finally, the log of per-capita income has a positive and statistically significant impact on flu incidence in the Google models, which is contrary to our expectations. Another potential explanation for the unexpected signs of these coefficients is that this specification—as mentioned in Section 2—is bias and inconsistent due the lagged dependent variable paired with regional fixed effects. The demographic variables are mostly insignificant across all specifications. This is likely because there is little variation in these variables over time.

6 Conclusion

This paper provides strong evidence that the spread of the influenza virus is linked with labor market conditions. Labor market-related activities like using public transportation and carpools, working in offices, and putting children in daycare are likely conducive to the spread of the flu. We hypothesize that increases in the numbers of people working or looking for work increase the spread of the flu. We test our theory using both Census Division-level data from the CDC and state-level data from Google Flu Trends. We find that a one percentage point increase in the percent of the population that is going to work (or actively looking for work) will increase the prevalence of the flu-related physician visits by a range of 8.1 to 9.6 flu-related physician visits per 1000 physician visits for all reasons. Out of every 100,000 new employees, we expect 33 additional flu-related doctor visits above the mean, or

a new total of 2,246 cases for the 100,000 individuals.

These results are relevant for several reasons. First, employment conditions can be forecasted, to a fairly accurate degree, several months in advance. This information could be used by the CDC to plan for the severity of an upcoming flu season. For example, if the economy is on an upswing, the CDC should plan for an above normal increase in flu incidence. These results also are of particular interest to firms concerning productivity during the flu season. Employers should consider that the lost productivity from many employees becoming infected with influenza is likely greater than the lost productivity from a few infected individuals taking sick leave. Workers concerned about missing pay or losing their jobs as the result of staying home from work due to illness will be less likely to heed early signs of influenza infection and stay home. Since a person may be infectious while experiencing mild symptoms, this greatly increases the probability that the virus will spread to other workers in the firm. This implies that firms should consider more generous "sick day" policies, particularly during the flu season.

Future research should concentrate on more micro-level data or testing this relationship internationally. Individual level data will allow a more direct estimation of the impact of employment status on the incidence of flu. Additionally, this will allow the determination of what types of employment are most likely to aid in the spread of the flu virus.

References

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*(375), 598–606.

Beggs, J. J. and M. Nerlove (1988). Biases in dynamic models with fixed effects. *Economics Letters 26*(1), 29 – 31.

Brankston, G., L. Gitterman, Z. Hirji, C. Lemieux, and M. Gardam (2007). Transmission of influenza a in human beings. *The Lancet Infectious Diseases 7*(4), 257 – 265.

Cannell, J. J., M. Zasloff, C. F. Garland, R. Scragg, and E. Giovannucci (2008). On the epidemiology of influenza. *Virology Journal 5*, 29.

CDC (2010, January). Seasonal influenza (flu). `http://www.cdc.gov/flu/`.

Cherry, D., E. Hing, D. Woodwell, and E. Rechtsteiner (2006). National ambulatory medical care survey: 2006 summary.

Cox, N. J. and K. Subbarao (2000). Global epidemiology of influenza: past and present. *Annual Review Of Medicine 51*, 407 – 421.

Gerdtham, U. and M. Johannesson (1999). New estimates of the demand for health: results based on a categorical health measure and swedish micro data. *Social Science & Medicine (1982) 49*(10), 1325 – 1332.

Gerdtham, U. and C. J. Ruhm (2006). Deaths rise in good economic times: evidence from the oecd. *Economics And Human Biology 4*(3), 298 – 316.

Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant (2009). Detecting influenza epidemics using search engine query data. *Nature 457*(7232), 1012 – 1014.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics*, pp. 221–233. Berkeley, Calif.: Univ. California Press.

Molinari, N.-A. M., I. R. Ortega-Sanchez, M. L. Messonnier, W. W. Thompson, P. M. Wortley, E. Weintraub, and C. B. Bridges (2007). The annual impact of seasonal influenza in the us: Measuring disease burden and costs. *Vaccine 25*(27), 5086 – 5096.

Nickell, S. J. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*(6), 1417 – 1426.

Ruhm, C. J. (2000). Are recessions good for your health? *Quarterly Journal of Economics 115*(2), 617 – 650.

Ruhm, C. J. (2003). Good times make you sick. *Journal Of Health Economics 22*(4), 637 – 658.

Ruhm, C. J. (2005). Healthy living in hard times. *Journal of Health Economics 24*(2), 341 – 363.

Ruhm, C. J. (2007). A healthy economy can break your heart. *Demography 44*(4), 829 – 848.

Ruhm, C. J. and W. E. Black (2002). Does drinking really decrease in bad times?. *Journal of Health Economics 21*(4), 659 – 678.

Tellier, R. (2006). Review of aerosol transmission of influenza a virus. *Emerging Infectious Diseases 12*(11), 1657 – 1662.

White, J. H. (2008). Vitamin d signaling, infectious diseases, and regulation of innate immunity. *Infection And Immunity 76*(9), 3837 – 3843.

Table 1: Summary Statistics

Variable Description	Mean	Std. Dev	Min	Max
CDC Flu Severity Variable	2.213	1.536	0.000	9.536
Google Flu Severity Variable	2.197	1.437	0.288	11.986
Employment Rate	63.006	2.235	57.766	70.255
% of Pop. in Labor Force	66.308	2.101	61.204	72.609
% of Pop. w/o Health Ins.	16.869	3.949	9.121	25.534
% of Pop. w/ Flu Shot	36.293	5.872	27.704	53.110
Log of Per-Capita Income	10.526	0.104	10.261	10.802
% Rural Population	19.837	8.584	7.214	42.918
% of Pop. Under 15	6.846	0.507	5.754	7.975
% of Pop. Over 65	12.506	1.091	10.840	13.855
% of Pop. w/ H.S. Degree	83.790	3.274	76.263	92.413
% of Pop. w/ B.A. Degree	26.683	2.859	19.402	34.485
Num. Obs	594			
First Flu Season	1997-1998			
Last Flu Season	2007-2008			

Table 2: FD2SLS Regression Models Using CDC Flu Data

	Employment Rate				Percent Labor Force			
	Base Model	Quadratic Month Trends	Februrary Peak Trends	Month Fixed Effects	Base Model	Quadratic Month Trends	Februrary Peak Trends	Month Fixed Effects
Employment Rate	0.732***	0.441***	0.365***	0.002				
	(0.09)	(0.08)	(0.11)	(0.06)				
% of Pop. in Labor Force					0.669***	0.483**	0.444***	0.261***
					(0.18)	(0.20)	(0.14)	(0.09)
Lagged Log of Flu Variable	2.042***	2.052***	1.154***	1.228***	1.979***	2.213***	1.172***	1.252***
	(0.05)	(0.22)	(0.12)	(0.14)	(0.05)	(0.20)	(0.12)	(0.14)
Log of Per-Capita Income	-20.343***	-19.759***	-18.138***	-19.956***	-15.493**	-16.874**	-16.023***	-19.609***
	(6.24)	(6.24)	(3.12)	(3.94)	(7.84)	(7.59)	(3.51)	(3.90)
N	594	594	594	594	594	594	594	594
F-Stat on Excl. Instruments	84.825	105.558	48.072	50.458	69.859	86.716	34.945	51.441
First Stage Partial R^2	0.101	0.052	0.154	0.126	0.100	0.039	0.123	0.122

Notes:
Robust standard errors clustered at the Census Division shown in parentheses.
All regressions run for the flu season spans runs from October to May.
All regressions are weighted by Census Division population.
* p<0.10, ** p<0.05, *** p<0.01

Table 3: FD2SLS Robustness Check Using Different Timing of Flu Variable

	Employment Rate			% Labor Force		
Timing of Flu Rates	Quad. Trend	Feb. Peak	Month FE	Quad. Trend	Feb. Peak	Month FE
Calendar Month	0.051	0.321**	-0.354***	-0.515***	0.136	-0.480***
Starting in Week 2	0.336**	0.405***	-0.173***	-0.041	0.325**	-0.171
Starting in Week 3	0.398***	0.385***	-0.067	0.207	0.405***	0.090
Starting in Week 4	0.441***	0.365***	0.002	0.483**	0.444***	0.261***
Starting in Week 5	0.252***	0.290***	0.030	0.342***	0.408***	0.277***
Starting in Week 6	0.254***	0.364***	0.066	0.519***	0.502***	0.348***

Notes:
Robust standard errors clustered at the Census Division level not shown.
All regressions run for the flu season which spans from October to May.
All regressions are weighted by Census Division population.
* p<0.10, ** p<0.05, *** p<0.01

Table 4: FD2SLS Regressions using Google Flu Data

	Employment Rate				Percent Labor Force			
	Base Model	Quadratic Month Trends	Februrary Peak Trends	Month Fixed Effects	Base Model	Quadratic Month Trends	Februrary Peak Trends	Month Fixed Effects
Employment Rate	0.459***	0.189***	0.307***	0.017				
	(0.04)	(0.04)	(0.05)	(0.04)				
% of Pop. in Labor Force					0.460***	0.107**	0.326***	0.019
					(0.08)	(0.05)	(0.07)	(0.05)
Lagged Log of Flu Variable	1.744***	0.979***	1.123***	1.028***	1.737***	1.028***	1.150***	1.028***
	(0.07)	(0.03)	(0.09)	(0.03)	(0.07)	(0.04)	(0.10)	(0.03)
Log of Per-Capita Income	-9.060***	-8.489***	-4.768*	-5.589*	-8.986***	-7.864***	-4.993*	-5.618*
	(3.22)	(2.89)	(2.85)	(3.08)	(3.41)	(2.87)	(2.92)	(3.05)
N	1836	1836	1836	1836	1836	1836	1836	1836
F-Stat on Excl. Instruments	165.395	206.304	778.214	300.754	172.010	181.350	395.301	292.902
First Stage Partial R^2	0.091	0.167	0.208	0.231	0.086	0.155	0.180	0.228

Notes:
Robust standard errors clustered at the state level shown in parentheses.
All regressions run for the flu season which spans from October to May.
All regressions are weighted by state population.
* p<0.10, ** p<0.05, *** p<0.01

Table 5: WOLS Regressions

	CDC Flu Rate		Google Flu Rate	
	Emp. Rate	% Labor Force	Emp. Rate	% Labor Force
Employment Rate	0.2243**		0.0205	
	(0.086)		(0.018)	
% of Pop. in Labor Force		0.2785**		0.1272***
		(0.116)		(0.038)
Lagged Log of Flu Variable	0.7672***	0.7340***	0.4346***	0.4455***
	(0.045)	(0.031)	(0.015)	(0.014)
% of Pop. w/o Health Ins.	-0.0929*	-0.1128*	-0.0144	-0.0102
	(0.048)	(0.056)	(0.027)	(0.025)
% of Pop. w/ Flu Shot	-0.0276	-0.0340	0.0143*	0.0098
	(0.021)	(0.031)	(0.007)	(0.007)
Log of Per-Capita Income	1.1963	2.7344	4.5608***	4.7517***
	(1.875)	(1.787)	(1.312)	(1.285)
% Rural Population	0.1048	-0.2500	0.1279	0.1108
	(0.242)	(0.269)	(0.174)	(0.175)
% of Pop. Under 15	0.1733	0.1085	-0.6314	-0.6603
	(0.167)	(0.213)	(0.759)	(0.767)
% of Pop. Over 65	-0.6997	-0.5528	0.0077	0.0301**
	(0.494)	(0.513)	(0.015)	(0.014)
% of Pop. w/ H.S. Degree	-0.0113	-0.0128	0.1481	0.1354*
	(0.077)	(0.091)	(0.091)	(0.072)
% of Pop. w/ B.A. Degree	-0.0208	0.0087	0.0114	0.0057
	(0.072)	(0.081)	(0.018)	(0.018)
N	594	594	1836	1836
R^2	0.469	0.468	0.268	0.277

Notes:
Robust standard errors clustered at the area level shown in parentheses.
All regressions are weighted by population.
All regressions run for the flu season which spans from October to May.
Models include region and flu season fixed-effects, and region-specific linear flu season trends.
* p<0.10, ** p<0.05, *** p<0.01