LUND UNIVERSITY
School of Economics and Management

Marijuana use and labor productivity in the United States

by

Bogdan Voronetskyy

August 2019

Master's Programme in Economics

Supervisor: Milda Norkute

Abstract

This thesis investigates the role of marijuana use in labor productivity in the United States with the aim to contribute to the discussion and understanding of economic consequences of marijuana legalization. The data is sourced from various federal agencies to create a panel of 50 U.S. states and the District of Columbia for the period of 2005-2017. A panel data model with fixed effects is estimated using instrumental variables to account for biases arising from endogeneity problems stemming from simultaneity and measurement errors. The results suggest that marijuana use does not have a statistically significant effect on labor productivity.

Keywords: labor productivity, drug use, marijuana, instrumental variables

Contents

1. Introduction

It is no secret that marijuana use and the extent of its legality in the United States is a topic of considerable debate. This comes as no surprise, since marijuana is the most used illicit drug in the United States, with lifetime use among individuals aged 26 and older reaching almost 50% in 2017 (NSDUH, 2017). Marijuana legalization is discussed by health activists, family values proponents and politicians who may seek to gain local government seats based on support or opposition towards marijuana by the general public in their state. Proponents point out the possible benefits of lower crime rates and tax income, while a more conservative side fears for a nation of drug addicts and marijuana being a gateway to heavier drugs. Even though marijuana is still prohibited under federal law, many states have independently passed laws to legalize it, and it has been shown that these states see an increase in marijuana consumption among adults following legalization (Sabia & Nguyen, 2016).

There are reasonable grounds for legalization concerns. Marijuana use has numerous reported health implications specifically affecting brain function. Heavy users may suffer from reduced attention, memory and learning functions which in some cases may be permanent, implying that beginning use at a young age may affect future education attainment and work performance (Batalla *et al.,* 2013; Filbey *et al.,* 2014). In a relevant study, Grant *et al.,* (2003) find that although there might be a decrease in the ability to learn and remember new information in chronic users, other cognitive abilities are unaffected. These health effects however, do not necessarily translate into users' subjective effects while being under marijuana influence. Users report that in addition to experiencing feelings of stress-relief, relaxation and well-being, while under the influence of marijuana they are more concentrated, creative, experience a rush of ideas and enhanced recollection, while learning itself is slower, the retention of learned material appears to be stronger (Osborne & Fogel, 2008). The above discrepancy between longer-term effects and perhaps more positive short-term effects is interesting from an economic point of view. As the majority of marijuana users in the U.S. fall into the working age population, and are employed, the effect that marijuana use has on their productivity is of interest to both employers, who are aware of this issue and have implemented workplace drug testing policies in response, and policy makers who are actively discussing further leniency towards marijuana (National Survey on Drug Use and Health, 2017).

In the public's eyes it is almost an axiom that marijuana use should negatively affect workforce productivity, however this notion is not backed by substantial research. There is a limited number of studies which provide inconclusive results. Previous research is largely dated and precludes the time of widespread legalization of marijuana in the United States. Moreover, many studies do not address possible endogeneity of marijuana use and other covariates.

This thesis aims to contribute to the understanding of the consequences of marijuana legalization by capturing the effect of marijuana use on labor productivity in the United States. This link is of interest since labor productivity is a fundamental economic indicator through which legalization, and subsequent increased marijuana use may affect broader economic outcomes. A panel data model with fixed effects is considered, which is estimated by employing instrumental variable estimation approach that allows to address not only endogeneity problems arising from unobserved state level time-invariant effects, but also from other sources such as simultaneity and measurement errors (Wooldridge, 2010). It is important to note that labor productivity is measured as average annual real wage in a state and marijuana use is defined as a percent of population that has consumed marijuana in any form in the past year. In addition, total marijuana use is further decomposed into age groups, allowing for the examination of not only net effects, but also effects of use among a specific age group, thereby providing a form of a robustness check. The data used is a panel of 50 U.S. states and the District of Columbia for 2005-2017 time period, the data is sourced from various federal agencies and statistical bureaus. Additional variables are used to control for other changes in labor productivity based on a model utilized by Carlino & Voith (1992) in their study on the sources of labor productivity differences across U.S. states. The results suggest that after controlling for endogeneity, contrary to public notion, marijuana use has no effect on aggregate labor productivity.

The paper is organized as follows. Section 2 contains the literature review. Section 3 outlines the theoretical basis for the analysis, describes the data used and states the empirical model. Section 4 presents the estimation results along with a short discussion. Section 5 is the conclusion.

2. Literature review

The literature on the subject of drug use and labor market outcomes is scarce and mainly deals with the effect of drug use on employment status and/or on wage differences. While there is a general consensus that use of drugs reduces the probability of being employed, there is mixed evidence on the effect drugs have, if any, on wages and by extension labor productivity.

Harwood *et al*. (1984) find a negative effect of marijuana use on household income, however their methodology was called into question by subsequent research. French *et al*., (1990) find that individuals who were admitted to rehabilitation saw an increase in their earnings after being discharged from the rehabilitation center. Assuming that drug use was reduced after rehab, this suggests that lower drug use increases earnings. On the contrary, Kaestner (1991) after controlling for self-selection finds that marijuana and cocaine users are expected to have 17%-19% higher wages than non-drug users. Similarly, Gill & Michaels, (1992) examine a range of both soft-drug and hard-drug users[1]. They find that after controlling for self-selection to use drugs and the decision to find a job, drug users while having lower employment levels than non-drug users, had higher wages once employed. Register & Williams (1992) use an instrumental variable (IV) approach relying on a single survey year of 1984 to estimate the effect of marijuana and cocaine use on labor productivity as measured by wages. Their findings suggest that while chronic use of marijuana has negative effects on productivity, the net productivity effect for all marijuana users is positive. Kandel *et al.,* (1995) using a sample of employed adult males find that drug use has positive effects on earnings in the early stages of employment (late-twenties), while it has negative effects in the later stages (mid-thirties). In a related study, DeSimone (2002) uses an IV approach to find that both cocaine and marijuana use is negatively related to the likelihood of employment, he however does not provide insight into wages.

[1] What exactly constitutes a soft or hard drug is dependent on legislation. Gill & Michaels (1992) treat amphetamines, marijuana and tranquilizers as soft drugs. While hard drugs include, Cocaine, Heroin, inhalants, psychedelics and others.

3. Theoretical and empirical framework

3.1 Theoretical model

Following the model utilized by Carlino & Voith (1992) to evaluate the sources of labor productivity differentials across U.S. states, I base my analysis on a constant elasticity of substitution (CES) production function which represents output in state *i* at time *t* (Q_{it}) as:

$$Q_{it} = A_{it} f(K_{it} L_{it}) \lambda^h \tag{1}$$

Since the above function assumes CES it can be written as:

$$Q_{it} = A_{it} \left[\delta K_{it}^{-h\rho} + (1 - \delta) L_{it}^{-h\rho} \right]^{-(1/\rho)} \tag{2}$$

Where K_{it} is the stock of capital, L_{it} is labor and A_{it} is total factor productivity.

Carlino & Voith (1992) then point out that in optimality labor is hired until its marginal product equals the wage rate. The marginal product of labor is given by:

$$W_{it} = \frac{\partial Q_{it}}{\partial L_{it}} = (1 - \delta) A_{it}^{-\rho} Q_{it}^{1+\rho} L_{it}^{(-1+h\rho)} \tag{3}$$

The main point being that states' labor productivity can be approximated by wages. Furthermore, as Carlino & Voith (1992) themselves follow Sveikauskas (1975), $f(\cdot)\lambda^h$ in Equation (1) is assumed to be identical across states. Therefore, differences in output should arise from differences in total factor productivity (A_{it}) which itself is a function of several key state characteristics with examples such as human capital, urbanization, infrastructure and industrial mix. Denoting the k-number of factors relevant for explaining differences in productivity across states as X_{kit}, total factor productivity can then be further decomposed to:

$$A_{it} = exp \left[a_0 + \sum_{k=1}^{k} a_k X_{kit} \right] \tag{4}$$

To get an empirical equation for estimating state labor productivity as wages we need to plug (4) into (3) and take logs of both sides:

$$\ln W_{it} = \beta_0 + \sum_{k=1}^{K} \beta_k X_{kit} + \theta \ln Q_{it} + \gamma \ln L_{it} \tag{5}$$

Now marijuana use can be included in Equation (5) as an additional explanatory variable:

$$\ln W_{it} = \beta_0 + \sum_{k=1}^{K} \beta_k X_{kit} + \theta \ln Q_{it} + \gamma \ln L_{it} + \omega M_{it} \tag{6}$$

Equation (6) is the model to be estimated by an appropriate method. Where $\ln W_{it}$ is the log of real wages, β_0 is the intercept, $\sum_{k=1}^{K} X_{kit}$ are the explanatory variables excluding labor and output, $\ln Q_{it}$ is the log of state output, $\ln L_{it}$ is the log of state labor M_{it} is marijuana use, and β_k, θ, γ, ω are the respective coefficients.

3.2 Data

Following the findings of Carlino & Voith (1992) on the sources of labor productivity differences across states, the logs of output, labor, and the explanatory variables forming X_{kit} in Equation (6) can be translated into the following observable variables in addition to marijuana use:

Table 3.2.1 Variables

Variable	*Unit*	*Time*	*Source*
Dependent			
Wage	Average annual wage in real terms (log form)	2005-2017	BLS
Independent			
Education	% of population with a bachelor's degree	2005-2017	ACS
Public road density	Miles of road per square mile of land area	2005-2017	FHA
Urbanization	% of population living in metropolitan areas	2000; 2010	Census
Gross State Product (GSP)	Total GSP in real terms (log form)	2005-2017	BEA
Labor	Number of employed people (log form)	2005-2017	BLS
Industrial mix	% contributed to GSP by an industry	2005-2017	BLS
Marijuana use	% of users in the past year	2005-2017	NSDUH

Abbreviations: BLS: Bureau of Labor Statistics; ACS: American Community Survey; FHA: Federal Highway Administration; BEA: Bureau of Economic Analysis; NSDUH: National Survey on Drug Use and Health provided by Substance Abuse and Mental Health Services Administration (SAMHSA).

Average annual real wage is a measure of labor productivity, converted to real terms using consumer price index in 1982-1984 dollars with 2012 as base year. Education as percent of people with a bachelor's degree is used as a proxy for human capital. Public road density per square mile of a state's land area is a proxy for infrastructure. This is useful since public roads are connecting or are directly a part of other infrastructure, for example public roads are built on the top of hydroelectric dams[2]. Public road density from here on is referred to as infrastructure for simplicity. Urbanization as percent of population living in metro areas is only available as part of the census which is conducted every ten years. Therefore, for this variable, values of 2000 are assumed for 2005-2009 and values of 2010 are assumed for 2010-2017. Both output Q_{it} and labor L_{it} in

[2] See Hoover Dam as a prime example.

Equation (6) are observable. Gross state product (GSP) measures the output and it is the state equivalent of GDP. Labor is defined as the amount of people currently employed. I refer to the variables in this paragraph, with the exception of wages, as main control variables for the sake of simplicity.

Industrial mix is defined as a percent contribution to real GSP made by a particular industry, it is important to stress that even though it is listed for convenience as a single variable in *Table 3.2.1*, it is in fact decomposed into 15 variables and every industry sector gets an estimated coefficient. To explain, say the total GSP in Alabama in 2005 was $1000, and the farming sector contributed $100 to that total, then we simply do $\frac{\$100}{\$1000} * 100 = 10\%$, this implies that the farming sector contributed 10% to GSP in that particular year in Alabama. After the above calculation is done for the 15 sectors, they together account for approximately 95% of each states' GSP, with the remaining 5% spread across a large number of smaller industry sectors. The industrial mix is included since different industry sectors have different effects on labor productivity (Carlino & Voith, 1992). Marijuana consumption may also have varying effects through different industries and it is important to control for this. Table 3.2.2 below presents the descriptive statistics for all the variables except marijuana, which is discussed separately further on. It also lists all the industry sectors[3].

[3] A more detailed description of industry sectors can be found in *Table 1.1* Appendix I

Table 3.2.2 Descriptive statistics

Variable	Mean	Min	Max	*Standard Deviation* Overall	Between	Within
Wage	46469	34267	86682	8578	8535	1437
Education	17.97%	10%	26%	2.90%	2.79%	0.86%
Infrastructure	2.07	0.02	22.15	2.98	3.01	0.04
Urbanization	73.39%	38.20%	100%	0.15	15.02%	1.18%
GSP	316300	26802	2587572	390874	392561	38364
Labor	2637883	254418	17000000	2862428	2882727	184427
Industries						
Agriculture	1.93%	0%	12.02%	2.29%	2.26%	0.46%
Oil	1.77%	0%	26.93%	4.28%	4.18%	1.10%
Mine	1.10%	0%	14.59%	2.28%	2.27%	0.39%
Construction	4.07%	0.91%	10.76%	1.17%	0.86%	0.80%
Durable	6.24%	0.06%	16.00%	3.40%	3.37%	0.66%
Non-durable	5.62%	0.11%	35.01%	3.69%	3.55%	1.10%
Wholesale	5.85%	0.85%	9.06%	1.49%	1.48%	0.29%
Retail	6.16%	1.04%	9.95%	1.31%	1.28%	0.35%
Transport	3.30%	0.33%	11.23%	1.60%	1.59%	0.26%
ICT	3.77%	1.11%	13.71%	1.84%	1.76%	0.57%
FIRE	18.45%	9.06%	44.62%	5.31%	5.28%	0.90%
PBS	10.58%	3.31%	25.63%	3.47%	3.44%	0.68%
EHS	8.75%	3.48%	13.91%	1.96%	1.90%	0.51%
Recreation	4.05%	2.24%	18.59%	2.20%	2.21%	0.24%
Government	14.02%	8.86%	36.02%	4.25%	4.21%	0.79%

From *Table 3.2.2* above we can make a few interesting observations. For one, the maximum value of urbanization is 100%. At first glance this seems strange but, it belongs to Washington D.C. which for data purposes is essentially treated as a city-state by U.S. statistical bureaus and is referred to as a state from here on for simplicity. For the same reason D.C. also has the highest value of infrastructure with 22 miles of public roads per square mile. Perhaps the most important observation we can make from the summary statistics is that the standard deviation between states is quite high, but within states the variables stay fairly constant. This may be problematic for a fixed-effects estimation since between effects are eliminated and all that is left is the within variation which may not be high enough to produce meaningful results.

Marijuana use is the variable of interest. It is expressed as a percentage share of total state population that has used marijuana in the past year and it is further decomposed into age groups in addition to state total use.

Table 3.2.3 Marijuana use descriptive statistics

		Age groups			
		Total (12+)	12-17	18-25	≥26
	Mean	12.16%	13.69%	30.89%	8.71%
	Min	5.30%	7.60%	1.60%	2.10%
	Max	26.51%	23%	53.20%	24.36%
Std. Deviation	Overall	3.53%	2.88%	6.93%	3.45%
	Between	3%	2.46%	6.20%	2.71%
	Within	1.89%	1.53%	3.21%	2.17%

The availability of data on marijuana use among different age groups can provide some robustness checks. For one, it is reasonable to expect that marijuana use among individuals aged 12-17 will not have an effect on productivity as measured by wages, since this age group is not substantially represented in the workforce and therefore their contribution to an average wage in a state should be very small. The same can be said about the 18-25 age group that consists of a large number of college and university students. However, the effects for this group can be more profound due to a larger share of internships, part-time jobs and military careers falling into this age group.

From *Table 3.2.3* it is evident that the 18-25 age group has by far the highest share of marijuana users in the 13 years under consideration, their mean is more than twice as large as the 12-17 age group which follows with 13.69%. Among the states[4], Washington D.C. and Alaska have the highest mean total marijuana consumption at 19.1% and 19% respectively. More generally, only 9 states out of 51 have mean total marijuana use less than 10% and 11 states break the 15% mark.

[4] Detailed state table is found in Appendix I *Table 1.2*

Lastly, once again the within variation in marijuana variables is approximately half of the between variation, with the exception of the ≥26 age group.

To sum up this section, the dataset is a strongly balanced panel with all the data sourced from different federal agencies. All variables except education, metropolitan population, GSP and labor required to be calculated manually, but to the best of my knowledge there are no mistakes in the final dataset.

3.3 Empirical specification

Equation (7) is the empirical continuation of Equation (6). It is a panel data model with fixed effects estimated using instrumental variable (IV) approach.

$$\ln W_{it} = \beta_{0i} + \sum_{k=1}^{k} \beta_k X_{it} + \theta \ln Q_{it} + \gamma \ln L_{it} + \varepsilon_{it} \tag{7}$$

It is a Two-Stage Least Squares (2SLS) estimator which is used to estimate the panel model with homogenous slopes. In the Data section it is mentioned that low within variation is problematic for a fixed-effects approach. However, a consistent estimation of panel data model with random effects requires to assume that the regressors are uncorrelated with the unobserved effects, an assumption which I believe is too strong to make in a case of state-level macro variables located in the same country. Therefore, a specification of panel data model with fixed effects is chosen despite the possibility of weak results.

While panel data model with fixed effects specification allows for endogeneity caused by the unobserved individual time-invariant effect, they do not address endogeneity caused by other sources. In the standard fixed effects models independent variables are assumed to be uncorrelated with the error term i.e. exogenous (Wooldridge, 2010, p.301). However, this assumption is not realistic given the variables used in this paper. The main example is marijuana use, when looking at the data a pattern emerges where states that have higher wages, also have higher marijuana use. This suggests simultaneity, where people with higher wages perhaps feel that they can afford to relax more and some of them will consume marijuana, resulting in increased use in their state. The same can be said about labor and industrial mix, since people might move to states where wages are higher and some companies might want to move to where wages are lower. More generally, it is too naive to assume that any variable used in this paper is determined exogenously without explicitly testing for endogeneity. Therefore, combining fixed effects with the instrumental variable approach allows to both remove unobserved individual specific effects and attempt to control for other sources of endogeneity. It is also useful to deal with measurement errors which are expected to be present since the data is survey based (Wooldridge, 2010, p.356). Especially with a variable of interest such as marijuana use which is reasonable to assume is underreported due to it being illegal in some states.

A good instrument satisfies two conditions. First it should be strong, meaning that it is highly correlated with the endogenous regressor and has a clear effect on it, this is a strong first stage. The second condition is validity, an instrument is valid when it affects the dependent variable only through the endogenous variable, otherwise the instrument would itself be correlated with the error term and hence suffer from endogeneity, this condition is also called the exclusion restriction (Angrist & Pischke, 2008, p.86).

Lags of endogenous variables are frequently used as instruments in the literature (see Hall, 1988; Yogo, 2004). The argument for doing so is usually as follows. Say that X_{it} is an endogenous explanatory variable. The argument is that since X_{it-1} precedes X_{it} in time, the instrument only affects Y_{it} through its effect on X_{it}, and since X_{it} is expected to be correlated with its own lag there is likely to be a strong first stage. Meaning that such an instrument is a good one, since it is both strong and valid. In practice more than one lag may be necessary for the above argument to hold, since only one lag may still directly affect the dependent variable, failing the exclusion restriction and invalidating the instrument. There is a tradeoff however, using longer lags as instruments means that the correlation between the endogenous variable and the lagged instrument may grow weaker the longer the lag length. Consequently, the weak instrument problem leads to biases in the two-stage least squares estimation. This tradeoff is important to consider since it has been shown that a strong instrument that is almost valid tends to bias 2SLS results only a little, while a weak instrument that is almost valid will bias the estimates much more severely (Hahn and Hausman, 2005 cited in Murray, 2006).

Diagnostic tests can help balance the tradeoff and decide on the number of lags to use as instruments. To get an idea of whether the instruments are strong enough one can use the Kleibergen-Paap LM statistic, which tests the null hypothesis that the instruments are not sufficiently correlated with the endogenous regressors. Moreover, this test does not need to assume that the errors are identically and independently distributed (i.i.d.), which is most likely not the case. In turn, the validity of the instruments can be tested with the Hansen J statistic, which tests the null that the excluded instruments are exogenous, and it also drops the i.i.d. assumption. Then by looking at the tests and the overall stability of the results one can decide on what number of lags to include as instruments. Of course, the tests should not be viewed as definite proof of a correct model specification, but as a guiding tool towards a reasonable one.

4. Results and discussion

4.1 Regression results and discussion

The following subsections present the results using three different estimation approaches. The estimators used are, OLS for baseline results, FE-IV with all variables instrumented, and finally FE-IV with only the endogenous variables instrumented. Each model is additionally reported with state and time specific fixed-effects in Appendix II. The reason for not including time FE specifications in the main text is that those models do not pass at least one of the necessary tests, they are however interesting for discussion purposes. Each table contains four regressions for total marijuana use and the three age groups. It is important to stress that age groups 12-17 and 18-25 are only included as a form of a robustness check, since it would be counterintuitive for those groups to produce significant results given the hypothesis of this thesis, as mentioned previously in the data section.

The endogenous variables for the second FE-IV specification are identified using the C-statistic which is reported instead of the Hausman endogeneity test when errors are not assumed to be i.i.d., in a case of i.i.d. errors they are asymptotically equivalent (Baum, Schaffer and Stillman, 2003). By estimating the models using different sets of instruments, I identified the two specifications that pass both the relevance and validity tests, that is Kleibergen-Paap LM and Hansen J tests respectively. Before presenting each model I more explicitly state the exact instruments used.

4.1.1 OLS Results

Below are the results obtained using the OLS estimator. These results show that the control variables are highly significant with the exception of some industries, and have the expected signs. Only urbanization has an unexpected negative sign, perhaps due to it capturing the negative effects of congestion. Marijuana use among total population and ≥26 age group appears to be positive and significant. Moreover, the statistical evidence that marijuana use affects productivity is stronger when time specific fixed effects are included in the model. Marijuana use among age groups 12-17 and 18-25 is insignificant as expected. Of course, those results are for comparison purposes only since OLS does not address endogeneity.

OLS no TE	(1)	(2)	(3)	(4)
VARIABLES	lwage	lwage	lwage	lwage
Education	0.553***	0.583***	0.572***	0.559***
	(0.0653)	(0.0633)	(0.0658)	(0.0649)
Infrastructure	0.0139***	0.0138***	0.0139***	0.0139***
	(0.000784)	(0.000813)	(0.000797)	(0.000780)
Urbanization	-0.152***	-0.155***	-0.153***	-0.153***
	(0.0151)	(0.0149)	(0.0163)	(0.0150)
lnGSP	0.828***	0.831***	0.829***	0.829***
	(0.0201)	(0.0201)	(0.0211)	(0.0200)
lnLabor	-0.805***	-0.810***	-0.807***	-0.807***
	(0.0209)	(0.0208)	(0.0218)	(0.0207)
Agricutlure	-0.971***	-0.970***	-0.967***	-0.977***
	(0.142)	(0.146)	(0.144)	(0.143)
Oil	-0.448***	-0.406***	-0.415***	-0.449***
	(0.116)	(0.118)	(0.118)	(0.115)
Mining	-0.0497	-0.0338	-0.0401	-0.0527
	(0.131)	(0.135)	(0.134)	(0.131)
Construction	1.342***	1.341***	1.335***	1.353***
	(0.207)	(0.206)	(0.207)	(0.208)
Durable manuf.	0.294**	0.340***	0.328**	0.296**
	(0.127)	(0.128)	(0.128)	(0.126)
Non-durable manuf.	-0.610***	-0.599***	-0.603***	-0.613***
	(0.120)	(0.121)	(0.121)	(0.120)
Wholesale trade	0.933***	0.906***	0.908***	0.928***
	(0.223)	(0.227)	(0.225)	(0.224)
Retail trade	-0.497**	-0.491**	-0.487**	-0.516***
	(0.197)	(0.199)	(0.199)	(0.197)
Transportation	-0.339***	-0.318**	-0.317**	-0.352***
	(0.129)	(0.134)	(0.133)	(0.129)
ICT	-0.152	-0.0920	-0.105	-0.149
	(0.140)	(0.137)	(0.139)	(0.139)
FIRE	-0.150	-0.136	-0.142	-0.152
	(0.102)	(0.105)	(0.105)	(0.102)
PBS	0.822***	0.866***	0.852***	0.822***
	(0.130)	(0.130)	(0.132)	(0.130)
EHS	0.696***	0.771***	0.739***	0.711***
	(0.139)	(0.137)	(0.146)	(0.135)
Recreation	0.156	0.213	0.197	0.155
	(0.144)	(0.143)	(0.145)	(0.144)
Government	-0.144	-0.129	-0.133	-0.145
	(0.134)	(0.138)	(0.137)	(0.133)
Marijuana (Total)	0.0957*			
	(0.0492)			
Marijuana (12-17)		0.00121		
		(0.0525)		
Marijuana (18-25)			0.0178	
			(0.0293)	
Marijuana (≥26)				0.0869*
				(0.0445)
Constant	12.06***	12.07***	12.06***	12.07***
	(0.106)	(0.109)	(0.109)	(0.106)
Time FE	NO	NO	NO	NO
Observations	663	663	663	663
R-squared	0.968	0.968	0.968	0.968

Robust standard errors in parentheses [*** p<0.01, ** p<0.05, * p<0.1]

4.1.2 FE-IV: Results with all regressors instrumented

This section presents the results for the panel data model with fixed effects which is estimated by instrumenting all the variables, referred to as Model 1. The control variables are instrumented with one lag, whereas every marijuana variable is instrumented with all three of its own lags.

In comparison to OLS, when controlling for endogeneity both the total and ≥26 marijuana variables become insignificant. Suggesting that OLS results are indeed misleading due to endogeneity. The coefficients associated with the main control variables of education, GSP and labor have the sign as expected and are statistically significant at 10% significance level. On the other hand, urbanization and infrastructure lost significance. This is perhaps the result of the aforementioned low within variation, particularly infrastructure varies extremely little, and urbanization is a discontinuous variable due to it being sourced from the census, meaning it changes value only once in the dataset.

With the addition of time fixed effects, the parameters associated with all the variables except agriculture are insignificant, this can be expected once again due to low within standard deviation. When time specific fixed effects are included, most of the within variation is now explained by the dummy variables, which leads to reduced significance of the remaining variables. The model passes the Hansen J test, but fails the Kleibergen-Paap test with a p-value close to one. The test results are presented below followed by regression results on the next page.

Table 4.1.2.1 Instrument relevance and validity tests for Model 1

	Time FE	
	NO	YES
Underidentification test (Kleibergen-Paap rk LM statistic):	9.441	0.483
Chi-sq(3) P-val =	0.024	0.9227
Hansen J statistic (overidentification test of all instruments):	0.055	1.58
Chi-sq(2) P-val =	0.9731	0.454

FE-IV Model 1 VARIABLES	(1) lwage	(2) lwage	(3) lwage	(4) lwage
Education	2.483**	3.484***	2.777***	2.490**
	(1.229)	(1.026)	(0.942)	(1.178)
Infrastructure	-0.0259	-0.0303	-0.0217	-0.0253
	(0.0294)	(0.0343)	(0.0300)	(0.0297)
Urbanization	-0.231	-0.283	-0.275	-0.228
	(0.234)	(0.283)	(0.240)	(0.236)
lnGSP	0.289**	0.281*	0.294**	0.295**
	(0.124)	(0.162)	(0.129)	(0.125)
lnLabor	-0.201*	-0.265*	-0.239**	-0.203*
	(0.115)	(0.141)	(0.113)	(0.112)
Agricutlure	-1.690**	-2.764***	-1.673**	-1.768**
	(0.744)	(0.893)	(0.714)	(0.739)
Oil	-1.099***	-1.070***	-1.172***	-1.096***
	(0.323)	(0.394)	(0.342)	(0.308)
Mining	-1.553***	-1.518***	-1.659***	-1.541***
	(0.461)	(0.581)	(0.494)	(0.446)
Construction	-0.326	-0.264	-0.364	-0.319
	(0.447)	(0.558)	(0.480)	(0.441)
Durable manuf.	-1.982***	-1.537*	-2.104***	-1.943***
	(0.632)	(0.886)	(0.681)	(0.627)
Non-durable manuf.	-0.542	-0.385	-0.575	-0.528
	(0.466)	(0.616)	(0.504)	(0.467)
Wholesale trade	0.453	0.288	0.328	0.443
	(0.867)	(1.024)	(0.868)	(0.852)
Retail trade	-2.444***	-2.072**	-2.407***	-2.416***
	(0.650)	(0.812)	(0.664)	(0.638)
Transportation	-0.609	-0.537	-0.663	-0.557
	(0.678)	(0.883)	(0.764)	(0.682)
ICT	-0.522	-0.577	-0.585	-0.540
	(0.483)	(0.583)	(0.498)	(0.466)
FIRE	-1.427**	-1.401*	-1.487**	-1.433**
	(0.586)	(0.724)	(0.614)	(0.557)
PBS	-0.826	-1.035	-0.929	-0.839
	(0.748)	(0.904)	(0.724)	(0.734)
EHS	0.0833	-0.0204	-0.0730	0.0948
	(1.084)	(1.464)	(1.144)	(1.050)
Recreation	-2.460***	-2.043**	-2.402***	-2.424***
	(0.868)	(1.016)	(0.876)	(0.864)
Government	-1.164*	-0.836	-1.322*	-1.078*
	(0.655)	(0.902)	(0.728)	(0.653)
Marijuana (Total)	0.101			
	(0.165)			
Marijuana (12-17)		-0.176		
		(0.148)		
Marijuana (18-25)			0.0232	
			(0.0720)	
Marijuana (≥26)				0.115
				(0.149)
Time FE	NO	NO	NO	NO
Observations	510	510	510	510
R-squared	0.710	0.603	0.686	0.710
Number of ID	51	51	51	51

Robust standard errors in parentheses [*** p<0.01, ** p<0.05, * p<0.1]

4.1.3 FE-IV: Results with only endogenous regressors instrumented

This model has only the endogenous variables instrumented, referred to as Model 2. Out of all the variables, only 10 out of 15 industries do not appear to be endogenous, namely agriculture, oil, durable goods manufacturing, transportation, ICT, FIRE, professional business services (pbs), educational and healthcare services, entertainment and government. Most of those are intuitive not to be determined endogenously, for example, irrespective of wages the oil industry cannot relocate. Government, education, healthcare, insurance, real estate and transportation also have to be present irrespective of wages.

The endogenous variables are instrumented with both 1st and 2nd lags in this specification, with some exceptions. Labor and wholesale trade are instrumented with 1 lag only, using more than 1 lag of those particular variables as instruments fails the Hansen J test. Marijuana is, just as in Model 1, instrumented with all lags up to and including the 3rd lag.

In contrast to Model 1, adding time fixed effects to this model increases the significance of most of the coefficients. Furthermore, coefficients associated with total marijuana and marijuana ≥26 also become highly significant. However, the time fixed effects specification passes the Hansen J test only at 5% significance level, and it fails the Kleibergen-Paap relevance of the instruments test. Once again, test results are presented below followed by regression results on the next page.

Table 4.1.3.1 Instrument relevance and validity tests for Model 2

	Time FE	
	NO	YES
Underidentification test (Kleibergen-Paap rk LM statistic):	20.525	11.166
Chi-sq(10) P-val =	0.0247	0.3447
Hansen J statistic (overidentification test of all instruments):	10.66	15.662
Chi-sq(9) P-val =	0.2998	0.0743

FE-IV Model 2 VARIABLES	(1) lwage	(2) lwage	(3) lwage	(4) lwage
Retail trade	-1.519**	-1.277*	-1.419**	-1.517**
	(0.634)	(0.694)	(0.658)	(0.634)
Non-durable manuf.	0.204	0.171	0.293	0.170
	(0.370)	(0.444)	(0.394)	(0.366)
Wholesale trade	1.645**	1.501*	1.752**	1.621**
	(0.817)	(0.910)	(0.820)	(0.822)
Mining	-0.317	-0.394	-0.291	-0.337
	(0.469)	(0.544)	(0.504)	(0.466)
Construction	0.820*	0.850*	0.937**	0.812*
	(0.435)	(0.495)	(0.451)	(0.437)
Education	2.273***	2.986***	2.479***	2.248***
	(0.827)	(0.662)	(0.651)	(0.844)
Infrastructure	-0.0336	-0.0249	-0.0306	-0.0309
	(0.0243)	(0.0273)	(0.0248)	(0.0244)
Urbanization	-0.0867	-0.0807	-0.103	-0.0670
	(0.208)	(0.237)	(0.219)	(0.213)
lnGSP	0.308***	0.299**	0.324***	0.301***
	(0.106)	(0.121)	(0.107)	(0.104)
lnLabor	-0.224**	-0.316***	-0.276***	-0.221**
	(0.105)	(0.113)	(0.100)	(0.105)
Agriculture	-1.116***	-1.351***	-1.142***	-1.122***
	(0.390)	(0.418)	(0.383)	(0.380)
Oil	-0.317	-0.405	-0.305	-0.328
	(0.266)	(0.301)	(0.276)	(0.264)
Durable manuf.	-0.746*	-0.800*	-0.707*	-0.762*
	(0.401)	(0.444)	(0.413)	(0.393)
Transportation	-0.424	-0.548	-0.421	-0.433
	(0.356)	(0.419)	(0.376)	(0.352)
ICT	0.304	0.152	0.357	0.269
	(0.405)	(0.453)	(0.407)	(0.395)
FIRE	-0.394	-0.518	-0.388	-0.411
	(0.416)	(0.461)	(0.421)	(0.412)
PBS	-0.269	-0.449	-0.216	-0.306
	(0.522)	(0.568)	(0.521)	(0.511)
EHS	1.079*	0.793	1.158*	0.971
	(0.637)	(0.721)	(0.667)	(0.614)
Recreation	-1.488**	-1.544**	-1.354**	-1.573**
	(0.644)	(0.700)	(0.659)	(0.650)
Government	-0.277	-0.446	-0.293	-0.265
	(0.393)	(0.435)	(0.403)	(0.392)
Marijuana (Total)	0.106			
	(0.128)			
Marijuana (12-17)		-0.137		
		(0.103)		
Marijuana (18-25)			0.0444	
			(0.0621)	
Marijuana (≥26)				0.132
				(0.121)
Time FE	NO	NO	NO	NO
Observations	510	510	510	510
R-squared	0.734	0.679	0.715	0.736
Number of panels	51	51	51	51

Robust standard errors in parentheses [*** p<0.01, ** p<0.05, * p<0.1]

4.2 Limitations

I believe the main limitation to be the nature of the data. First of all, measuring marijuana use as a percent of state population is not optimal for the purpose of studying its effect on labor productivity. Having a certain percent of population using marijuana does not guarantee that the overall consumption is higher or lower. Because of this it would be much more meaningful to define marijuana use by weight consumed in state i during time t. A 10% of users in one state may actually consume more by weight than 15% of users in another state. Unfortunately, to the best of my knowledge there is no such data available as of the time of writing. Another data limitation is that the NSDUH survey has quite large confidence intervals on marijuana use percentages, approximately 2%. Which is far from ideal considering some states have roughly 8% users and it is a big difference between having 6% and 10% of population use marijuana.

Another implicit limitation is the assumption of homogenous effects of marijuana on labor productivity. It is reasonable to think that marijuana use has different effects in republican, religious, family-value states such as Alabama, and more progressive states such as California and Washington.

This study also fails to distinguish between chronic and casual users of marijuana. Labor productivity outcomes for those groups are probably vastly different. As for example, French *et al.* (2001) find that non-chronic drug use did not affect employment status, arguing that drug testing prior to hiring should focus on chronic users instead of casual users.

Finally, the choice of using lags of endogenous variables as instruments while common, is not substantially explored from a theoretical standpoint. Certainly, superior instruments for marijuana use can be found. However, identifying and arguing for the validity of an instrument candidate is an immense task and due to time limitations this is outside the scope of this study.

5. Conclusion

This study examined whether marijuana use has an effect on labor productivity aiming to contribute to the knowledge on the relationship between drug use and labor market outcomes. When one controls for endogeneity, the results suggest that marijuana use plays no significant role in determining labor productivity in the United States since both FE-IV specifications that satisfy the necessary tests show marijuana to be insignificant. Additionally, it is very likely that marijuana is endogenous to wages and future studies on this topic may need to address this before making inference.

While *Model 2 with Time Fixed-Effects* shows significance for marijuana variables, it cannot be presented and relied upon as a main result since it fails the strength test and is therefore biased to some degree due to weak instruments. Nevertheless, panel data model specifications with state and time specific fixed effects are interesting for comparison and transparency reasons. However, I have not found a model specification that includes time fixed-effects and passes the Kleibergen-Paap LM and Hansen J tests, although *Model 2 with TFE* comes somewhat close.

Even though public opinion on the matter of marijuana use is generally negative, it appears that from a purely economic standpoint things are not as clear-cut as they are thought to be. The results are useful to add to the discussion and understanding of legalization and its consequences. However, they may not be generalizable to countries other than the United States due the data used. In the end, such a research question is held back by data limitations and lack of natural experiments that warrant better research designs. Future research can substantially improve upon the findings by addressing the limitations listed in the previous section.

References

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Baum, C.F., Schaffer, M.E. and Stillman, S., 2003. Instrumental variables and GMM: Estimation and testing. *The Stata Journal*, *3*(1), pp.1-31.

Batalla, A., Bhattacharyya, S., Yuecel, M., Fusar-Poli, P., Crippa, J.A., Nogue, S., Torrens, M., Pujol, J., Farre, M. and Martin-Santos, R., 2013. Structural and functional imaging studies in chronic cannabis users: a systematic review of adolescent and adult findings. *PloS one*, *8*(2), p.e55821.

Carlino, G. A., & Voith, R. (1992). Accounting for differences in aggregate state productivity. *Regional Science and Urban Economics*, *22*(4), 597-617.

DeSimone, J. (2002). Illegal drug use and employment. *Journal of Labor Economics*, *20*(4), 952-977.

Filbey, F.M., Aslan, S., Calhoun, V.D., Spence, J.S., Damaraju, E., Caprihan, A. and Segall, J., 2014. Long-term effects of marijuana use on the brain. *Proceedings of the National Academy of Sciences*, *111*(47), pp.16913-16918.

French, M. T., Zarkin, G. A., Hubbard, R. L., & Rachal, J. V. (1991). The impact of time in treatment on the employment and earnings of drug abusers. *American Journal of Public Health*, *81*(7), 904-907.

French, M. T., Roebuck, M. C., & Alexandre, P. K. (2001). Illicit drug use, employment, and labor force participation. *Southern Economic Journal*, 349-368.

Grant, I., Gonzalez, R., Carey, C. L., Natarajan, L., & Wolfson, T. (2003). Non-acute (residual) neurocognitive effects of cannabis use: a meta-analytic study. *Journal of the International Neuropsychological Society*, *9*(5), 679-689.

Gill, A. M., & Michaels, R. J. (1992). Does drug use lower wages?. *ILR Review*, *45*(3), 419-434.

Hall, R. E. (1988). Intertemporal substitution in consumption. *Journal of political economy*, *96*(2), 339-357.

Harwood, H. J., Napolitano, D. M., Kristiansen, P. L., & Collins, J. J. (1984). Economic costs to society of alcohol and drug abuse and mental illness: 1980. *Research Triangle Park, NC: Research Triangle Institute*, 00-01.

Kaestner, R. (1991). The effect of illicit drug use on the wages of young adults. *Journal of Labor Economics*, *9*(4), 381-412.

Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of economic Perspectives*, *20*(4), 111-132.

National Survey on Drug Use and Health. (2017). [pdf] Available at: https://www.samhsa.gov/data/nsduh/reports-detailed-tables-2017-NSDUH [Accessed 18 June 2019]

Register, C. A., & Williams, D. R. (1992). Labor market effects of marijuana and cocaine use among young men. *ILR Review*, *45*(3), 435-448.

Sabia, J.J. and Nguyen, T.T., 2018. The effect of medical marijuana laws on labor market outcomes. *The Journal of Law and Economics*, *61*(3), pp.361-396.

Sveikauskas, L. (1975). The productivity of cities. *The Quarterly Journal of Economics*, *89*(3), 393-413.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Yogo, M. (2004). Estimating the elasticity of intertemporal substitution when instruments are weak. *Review of Economics and Statistics*, *86*(3), 797-810.

Appendix I: Data

Table 1.1 Industrial mix composition

Industry	NAICS code
Agriculture, forestry, fishing, and hunting	11
Oil and gas extraction	2111
Mining (except oil and gas)	2121-2123
Construction	23
Durable goods manufacturing	31-33
Non-durable goods manufacturing	31-33
Wholesale trade	42
Retail trade	44-45
Transportation and warehousing	48-49
Information and communication technology (ICT)	51
Finance, insurance, real estate, rental, and leasing (FIRE)	52-53
Professional and business services (PBS)	54
Educational services, health care, and social assistance (EHS)	61-62
Arts, entertainment, recreation, accommodation, and food services	71-72
Government and government enterprises	92

Table 1.2 Mean marijuana use by age group and mean wages by state

State	Wage	Mean marijuana use by age group Total (12+)	12-17	18-25	≥26
Alabama	41990	8.9%	10.2%	21.6%	6.3%
Alaska	50141	19.0%	16.4%	36.7%	16.0%
Arizona	45760	11.7%	14.6%	26.8%	8.6%
Arkansas	38313	10.4%	12.0%	26.5%	7.5%
California	56900	13.8%	14.2%	32.0%	10.4%
Colorado	50692	17.7%	18.4%	40.2%	13.6%
Connecticut	62499	13.7%	16.5%	39.3%	9.3%
Delaware	51757	12.4%	14.8%	35.6%	8.2%
D.C.	82878	19.1%	16.4%	40.2%	14.8%
Florida	43850	11.2%	13.3%	30.6%	8.1%
Georgia	46662	10.8%	11.7%	27.1%	7.9%
Hawaii	44066	12.4%	15.2%	28.3%	9.6%
Idaho	36906	10.5%	12.9%	25.3%	7.5%
Illinois	52535	11.1%	12.6%	30.2%	7.6%
Indiana	41629	11.0%	12.5%	29.8%	7.4%
Iowa	40603	8.5%	11.6%	24.3%	5.2%
Kansas	41179	9.8%	12.2%	26.2%	6.4%
Kentucky	40774	10.1%	11.6%	27.0%	7.1%
Louisiana	43016	9.7%	10.5%	25.6%	6.6%
Maine	39290	16.0%	16.1%	40.8%	12.5%
Maryland	53962	11.1%	14.2%	33.5%	7.0%
Massachusetts	61601	15.3%	16.9%	41.6%	10.6%
Michigan	47578	13.8%	14.4%	33.5%	10.3%
Minnesota	49676	11.7%	12.1%	31.8%	8.3%
Mississippi	35936	8.5%	9.7%	22.5%	5.7%
Missouri	43001	11.0%	12.8%	28.8%	7.8%
Montana	36797	15.1%	17.2%	36.4%	11.2%
Nebraska	39669	9.5%	11.2%	26.7%	6.1%

Nevada	44861	12.9%	15.4%	28.9%	10.1%
New Hampshire	49012	15.3%	16.9%	42.4%	10.9%
New Jersey	59167	10.2%	12.3%	30.9%	6.9%
New Mexico	40475	13.2%	18.2%	31.9%	9.1%
New York	63695	13.3%	13.9%	34.2%	9.6%
North Carolina	43481	10.7%	12.5%	29.2%	7.4%
North Dakota	42424	8.7%	10.7%	24.3%	5.0%
Ohio	44412	11.6%	13.0%	30.5%	8.3%
Oklahoma	40712	10.2%	11.9%	24.8%	7.3%
Oregon	44664	18.0%	17.9%	38.6%	14.8%
Pennsylvania	48476	11.1%	12.9%	31.1%	7.6%
Rhode Island	47163	17.8%	18.2%	41.5%	12.9%
South Carolina	39604	10.3%	11.9%	27.8%	7.0%
South Dakota	36585	8.9%	11.5%	25.0%	5.7%
Tennessee	43743	10.4%	11.3%	26.6%	7.7%
Texas	50122	9.1%	11.3%	23.7%	6.1%
Utah	41432	8.0%	9.4%	18.1%	5.3%
Vermont	41534	18.4%	18.7%	46.9%	13.5%
Virginia	51454	10.5%	13.0%	29.5%	7.0%
Washington	52037	15.9%	14.8%	35.6%	12.8%
West Virginia	39198	10.2%	12.1%	28.4%	7.3%
Wisconsin	42471	10.9%	14.1%	29.9%	7.2%
Wyoming	43523	10.4%	14.0%	27.0%	7.1%
Total	46469	12.2%	13.7%	30.9%	8.7%

Appendix II: Estimation results of panel data models with state and time specific fixed effects

Tables begin on the next page.

OLS with time dummies	(1)	(2)	(3)	(4)
VARIABLES	lwage	lwage	lwage	lwage
Education	0.521***	0.543***	0.554***	0.523***
	(0.0549)	(0.0543)	(0.0559)	(0.0544)
Infrastructure	0.00862***	0.00850***	0.00840***	0.00858***
	(0.000854)	(0.000885)	(0.000872)	(0.000846)
Urbanization	-0.134***	-0.137***	-0.139***	-0.133***
	(0.0125)	(0.0121)	(0.0134)	(0.0124)
lnGSP	0.858***	0.863***	0.865***	0.858***
	(0.0161)	(0.0159)	(0.0170)	(0.0160)
lnLabor	-0.838***	-0.843***	-0.846***	-0.837***
	(0.0169)	(0.0165)	(0.0177)	(0.0168)
Agricutlure	-1.767***	-1.774***	-1.781***	-1.774***
	(0.145)	(0.147)	(0.147)	(0.146)
Oil	-1.418***	-1.390***	-1.385***	-1.431***
	(0.126)	(0.126)	(0.125)	(0.126)
Mining	-0.949***	-0.944***	-0.941***	-0.959***
	(0.134)	(0.137)	(0.135)	(0.135)
Construction	-0.669**	-0.654**	-0.658**	-0.685***
	(0.263)	(0.263)	(0.264)	(0.263)
Durable manuf.	-0.515***	-0.478***	-0.469***	-0.530***
	(0.130)	(0.130)	(0.129)	(0.131)
Non-durable manuf.	-1.491***	-1.486***	-1.486***	-1.501***
	(0.121)	(0.121)	(0.121)	(0.121)
Wholesale trade	0.0365	0.0110	0.00583	0.0319
	(0.202)	(0.203)	(0.203)	(0.202)
Retail trade	-1.511***	-1.529***	-1.543***	-1.532***
	(0.186)	(0.187)	(0.190)	(0.185)
Transportation	-1.213***	-1.199***	-1.206***	-1.241***
	(0.137)	(0.140)	(0.141)	(0.136)
ICT	-1.095***	-1.042***	-1.031***	-1.113***
	(0.133)	(0.132)	(0.131)	(0.133)
FIRE	-1.033***	-1.026***	-1.025***	-1.043***
	(0.113)	(0.114)	(0.114)	(0.113)
PBS	0.0723	0.102	0.112	0.0605
	(0.129)	(0.131)	(0.130)	(0.129)
EHS	0.00305	0.0738	0.106	-0.00535
	(0.130)	(0.132)	(0.134)	(0.128)
Recreation	-0.530***	-0.485***	-0.471***	-0.548***
	(0.130)	(0.131)	(0.129)	(0.130)
Government	-0.995***	-0.980***	-0.978***	-1.008***
	(0.135)	(0.137)	(0.136)	(0.135)
Marijuana (Total)	0.0926**			
	(0.0419)			
Marijuana (12-17)		0.00344		
		(0.0478)		
Marijuana (18-25)			-0.0172	
			(0.0256)	
Marijuana (≥26)				0.112***
				(0.0395)
Constant	13.04***	13.05***	13.07***	13.05***
	(0.121)	(0.120)	(0.123)	(0.121)
Time FE	YES	YES	YES	YES
Observations	663	663	663	663
R-squared	0.978	0.978	0.978	0.978

Robust standard errors in parentheses [*** p<0.01, ** p<0.05, * p<0.1]

FE-IV Model 1 with TFE	(1)	(2)	(3)	(4)
VARIABLES	lwage	lwage	lwage	lwage
Education	-2.940	0.273	4.007	5.467
	(6.314)	(2.109)	(2.731)	(12.62)
Infrastructure	-0.0532	-0.0584***	-0.0740**	-0.0771
	(0.0352)	(0.0155)	(0.0370)	(0.0617)
Urbanization	0.131	0.0800	0.0440	0.0455
	(0.261)	(0.172)	(0.238)	(0.332)
lnGSP	0.332	0.467***	0.762***	0.890
	(0.451)	(0.168)	(0.232)	(0.834)
lnLabor	-0.208	-0.403**	-0.716***	-0.845
	(0.496)	(0.178)	(0.250)	(0.946)
Agricutlure	-2.856**	-2.642***	-2.136**	-1.894
	(1.146)	(0.843)	(0.933)	(1.930)
Oil	-0.393	-0.891*	-1.503***	-1.685
	(1.047)	(0.521)	(0.544)	(2.054)
Mining	-0.274	-1.033	-1.910**	-2.162
	(1.570)	(0.767)	(0.829)	(3.024)
Construction	0.717	-0.119	-1.279	-1.529
	(1.831)	(0.838)	(1.020)	(3.609)
Durable manuf.	0.417	-0.319	-1.217	-1.471
	(1.613)	(0.911)	(0.907)	(3.048)
Non-durable manuf.	-0.535	-0.862	-1.269*	-1.240
	(0.898)	(0.570)	(0.654)	(1.461)
Wholesale trade	0.942	0.704	0.691	0.918
	(0.891)	(0.705)	(0.815)	(1.018)
Retail trade	-0.441	-0.812	-1.511	-1.387
	(1.173)	(0.832)	(0.939)	(1.876)
Transportation	-1.878	-0.819	0.644	1.623
	(2.811)	(0.895)	(1.249)	(5.179)
ICT	0.185	-0.227	-0.657	-0.704
	(0.936)	(0.644)	(0.710)	(1.637)
FIRE	-0.355	-0.545	-0.643	-0.539
	(0.739)	(0.565)	(0.630)	(0.860)
PBS	-0.399	-0.587	-0.653	-0.570
	(0.684)	(0.481)	(0.672)	(0.905)
EHS	-1.203	-0.0208	2.196	3.252
	(3.671)	(1.236)	(1.771)	(6.522)
Recreation	0.175	-0.245	-0.966	-0.866
	(1.292)	(0.881)	(1.124)	(2.164)
Government	0.213	-0.770	-1.632	-1.990
	(1.812)	(0.899)	(1.031)	(3.571)
Marijuana (Total)	0.145			
	(0.169)			
Marijuana (12-17)		-0.0265		
		(0.119)		
Marijuana (18-25)			0.0774	
			(0.120)	
Marijuana (≥26)				0.0762
				(0.218)
Time FE	YES	YES	YES	YES
Observations	510	510	510	510
R-squared	0.646	0.849	0.633	0.401
Number of panels	51	51	51	51

Robust standard errors in parentheses [*** p<0.01, ** p<0.05, * p<0.1]

FE-IV Model 2 with TFE	(1)	(2)	(3)	(4)
VARIABLES	lwage	lwage	lwage	lwage
Retail trade	-1.624***	-1.523***	-1.471***	-1.509***
	(0.385)	(0.420)	(0.412)	(0.389)
Non-durable manuf.	-1.086***	-1.045***	-1.008***	-1.075***
	(0.293)	(0.299)	(0.284)	(0.294)
Wholesale trade	0.713	0.557	0.611	0.792*
	(0.448)	(0.499)	(0.478)	(0.446)
Mining	-1.329***	-1.353***	-1.291***	-1.351***
	(0.315)	(0.339)	(0.339)	(0.311)
Construction	-0.351	-0.223	-0.270	-0.367
	(0.345)	(0.370)	(0.376)	(0.343)
Education	0.322	0.720	0.842	0.494
	(0.692)	(0.596)	(0.638)	(0.700)
Infrastructure	-0.0629***	-0.0579***	-0.0594***	-0.0579***
	(0.0142)	(0.0139)	(0.0143)	(0.0141)
Urbanization	0.0866	0.0871	0.0387	0.103
	(0.119)	(0.120)	(0.116)	(0.117)
lnGSP	0.656***	0.590***	0.638***	0.679***
	(0.0882)	(0.0873)	(0.0871)	(0.0867)
lnLabor	-0.506***	-0.526***	-0.561***	-0.536***
	(0.104)	(0.0989)	(0.102)	(0.105)
Agriculture	-1.209***	-1.385***	-1.349***	-1.186***
	(0.167)	(0.183)	(0.182)	(0.172)
Oil	-1.148***	-1.171***	-1.165***	-1.146***
	(0.201)	(0.209)	(0.204)	(0.202)
Durable manuf.	-1.081***	-1.145***	-1.095***	-1.057***
	(0.235)	(0.250)	(0.243)	(0.236)
Transportation	-0.385	-0.533**	-0.450*	-0.302
	(0.255)	(0.247)	(0.244)	(0.256)
ICT	-0.549**	-0.546**	-0.538**	-0.554**
	(0.251)	(0.267)	(0.253)	(0.250)
FIRE	-1.047***	-1.108***	-1.076***	-1.027***
	(0.227)	(0.241)	(0.240)	(0.228)
PBS	-0.175	-0.327	-0.237	-0.125
	(0.297)	(0.314)	(0.301)	(0.296)
EHS	1.013*	0.646	0.924*	1.075**
	(0.524)	(0.531)	(0.516)	(0.511)
Recreation	-1.018***	-0.941***	-0.922***	-1.120***
	(0.313)	(0.310)	(0.334)	(0.333)
Government	-0.521**	-0.800***	-0.713***	-0.492**
	(0.251)	(0.258)	(0.251)	(0.249)
Marijuana (Total)	0.192***			
	(0.0634)			
Marijuana (12-17)		0.00962		
		(0.0632)		
Marijuana (18-25)			-0.00219	
			(0.0435)	
Marijuana (≥26)				0.220***
				(0.0612)
Time FE	YES	YES	YES	YES
Observations	510	510	510	510
R-squared	0.907	0.901	0.901	0.906
Number of ID	51	51	51	51

Robust standard errors in parentheses [*** p<0.01, ** p<0.05, * p<0.1]