

1 Temperature and Economic Outcomes

1.1 Temperature Aggregation

See the do-file "1_1_temperature_aggregation.do".

1.2 US Climate Impacts: County-Year Damages

See the do-file "1_2_climate_impacts_regressions.do" and R script "create_figures_tables.R".

1.

Figure 1 displays the linear relationship between log transformed `emp_farm` and the vector of binned temperature controls, including county and year fixed effects. That is, it displays the estimates $\hat{\beta}$ obtained via OLS from the following regression on log farm employment in county i and year t :¹

$$\log(emp_farm)_{it} = X'_{it}\beta + \alpha_i + \gamma_t + \epsilon_{it},$$

The omitted temperature bin is 20-24°C. Thus, the coefficient of roughly 0.0015 on the $> 32^\circ\text{C}$ bin can be interpreted as "Relative to an additional 20-24°C day during the year, an additional $> 32^\circ\text{C}$ day is associated with an increase in farm employment of roughly 0.15%. The relationship is statistically significant at the 5% level."

2.

Figure 2 displays the relationship between log(per capita farm prop income) and temperature using the restricted cubic spline, including county and year fixed effects. Unfortunately, I did not have time to plot the marginal effects in Stata, where the marginal effects are the partial derivative of the log(per capita farm prop income) with respect to the average temperature within a given county.

3.

Table 1 displays the results of using the binned temperature estimator to design a test for whether we observe treatment effect heterogeneity. Specifically, I add interactions for the number of $> 32^\circ\text{C}$ days with the respective temperature bins. As in Exercise 1.1, the omitted bin is 20-24°C. While the majority of the estimated coefficients on the interaction terms are not statistically significant at the 5% level, the coefficients on `tempB0 × tempA32`

1. Note that $\log(emp_farm)_{it}$ is missing for the 1,624 observations with emp_farm_{it} equal to zero. These 1,624 observations comprise roughly 1.26% of the analysis sample.

and `temp4to8` \times `tempA32` are statistically significant at the 5% level. Conducting an F-test on the interaction terms jointly equalling zero yields an F-statistic of roughly 13; hence, we can reject the null hypothesis of no treatment effect heterogeneity.

2 Hedonic Air Quality Analysis

See the do-file “2_hedonic_air_quality_analysis.do” and R script “create_figures_tables.R”. To maintain the same sample throughout the questions in this section, I drop the 46 observations that were missing any of the variables included in the analysis, leaving me with 967 observations. Unless stated otherwise, “controls” refers to all “other relevant variables” listed in the data notes, including the three economic shock variables references in problem 1.

1.

Columns 1 and 2 of Table 2 displays the estimates obtained via OLS from regressing housing prices on pollution levels with and without controls. Without controls, my estimates imply that a one unit increase in the change in the annual geometric mean of total suspended particulates pollution (TSPs) from 1969-72 to 1977-80 is associated with roughly a 0.26% increase in housing values from 1970 to 1980 (statistically sig. at 5%). However, the point estimate is reduced by almost an order of magnitude after controlling for other housing price determinants and economic shocks and is no longer statistically significant at conventional levels. This difference implies that the regression without controls suffers from omitted variables bias, where the change in pollution is correlated with other housing price determinants.

There are many potential omitted variables biases stories that can be told here. For instance, increases in income could be positively correlated with increases in pollution if the former is pollution and economic activity are positively correlated. Increases in per-capita income are also likely linked to increases in housing values. Indeed, I find in the data that both sets of simple correlations - pollution and income, and housing values and income - are positive. These relationships are also positive (and statistically significant at the 5% level) when controlling for the other two economic shocks listed above (Columns 3 and 4 of Table 2), although the point estimates are close to zero.

2.

Two main assumptions must be satisfied for mid-decade regulatory status `tsp7576` to be a valid instrument for pollution changes `dgtsp`:

1. Relevance condition: the instrument must be correlated with the endogenous variable of interest, conditional on the other covariates. This assumption can be checked in the first-stage regression of 2SLS, the results of which are reported in the next section.
2. Exclusion restriction: the instrument is uncorrelated with the error term, conditional on the other covariates. While this assumption is intrinsically impossible to check, we can assess its plausibility by investigating whether the instrument is correlated with other observed determinants of the outcome of interest (log housing prices). Looking at the coefficients on the economic shock measures in Column 2 of Table 3, `dincome` is statistically significantly related to `tsp7576` at the 5% level while `dunemp` and `dmnfcg` are not. Although we are able to control for `dincome` in our regression, this finding is not reassuring that there are no unobserved determinants of the outcome of interest that are correlated with the instrument.

3.

Table 4 displays the results for this problem. Columns 1 and 2 display the first-stage relationship between regulation and air pollution changes without and with controls, respectively. Both with and without controls, the first stage relationship appears to be strong (i.e., EPA regulation appears to be a strong and statistically significant predictor of air pollution changes), though is slightly attenuated when including the controls. Interpreting our findings without (with) controls in Column 1 (Column 2), counties that were regulated in either 1975 or 1976 experienced a roughly 9.84 (8.05) unit reduction in the annual geometric mean of TSPs from 1969-72 to 1977-80 compared to counties that were not regulated (holding fixed the other covariates in the regression).

Columns 3 and 4 display the reduced form relationship between regulation and housing price changes without and with controls, respectively. Both coefficients are similar in magnitude and are statistically significant at the 5% level. Interpreting our findings without (with) controls in Column 3 (Column 4), counties that were regulated in either 1975 or 1976 experienced a roughly 4.45% (3.96%) increase in housing values from 1970 to 1980 compared to counties that were not regulated (holding fixed the other covariates in the regression).

Columns 5 and 6 display the 2SLS relationship between air pollution changes and housing price changes without and with controls, respectively. The 2SLS point estimates can be obtained by dividing the reduced form coefficient estimates by the first stage coefficient estimates. Both coefficients are similar in magnitude and are statistically significant at the 5% level. Interpreting our findings without (with) controls in Column 5 (Column 6), a 1 unit increase in TSPs from 1969-72 to 1977-80 is associated with roughly a 0.45% (0.49%) decrease in housing values from 1970 to 1980 (holding fixed the other covariates in the regression). Assuming the IV assumptions are met, these estimates reflect the local average treatment effect (LATE), or the average treatment effect for the subset of counties whose air pollution changed as a result of EPA regulation. If we had homogeneous treatment effects, then the LATE would equal the average treatment effect (ATE) by definition.

4.

Table 5 displays the results for this problem. All of the following results (first stage, second stage, reduced form) have the same sign as those obtained in the previous part.

Columns 1 and 2 display the first-stage relationship between the annual geometric mean of TSPs in 1974, `mtspgm74`, and air pollution changes, `dgtsp`, without and with controls, respectively. In both specifications, the first stage relationship appears to be strong, though is slightly attenuated when including the controls. Interpreting our findings without (with) controls, a one unit increase in annual geometric mean of TSPs in 1974 is associated with roughly 0.27 (0.24) unit reduction in the annual geometric mean of TSPs from 1969-72 to 1977-80 (holding fixed the other covariates in the regression).

Columns 3 and 4 display the reduced form relationship between our new instrumental variable and housing price changes without and with controls, respectively. The two estimates are nearly identical, though only the one with controls is statistically significant at the 5% level. Interpreting our findings without (with) controls, a one unit increase in an-

nual geometric mean of TSPs in 1974 is associated with roughly 0.12% (0.12%) increase in housing values from 1970 to 1980 (holding fixed the other covariates in the regression).

Columns 5 and 6 display the 2SLS relationship between air pollution changes and housing price changes without and with controls, respectively. As in the reduced form relationship, both point estimates are very similar, though only the one with controls is statistically significant at the 5% level. Interpreting our findings without (with), a one unit increase in TSPs from 1969-72 to 1977-80 is associated with roughly a 0.46% (0.51%) decrease in housing values from 1970 to 1980 (holding fixed the other covariates in the regression).

5.

Using this discontinuity in treatment assignment, we could derive an alternate estimator using a fuzzy RDD, where the probability of a county being regulated jumps discontinuously at the 75 unit cutoff value of 1974 TSPs. The reason we cannot use a sharp RDD is because of imperfect compliance with treatment assignment at the cutoff (i.e., some counties that are below the cutoff value are nevertheless regulated). The key identifying assumption in this design is continuity, which means that the expected potential outcomes are continuous at the cutoff.

There are a few ways this identifying assumption can fail. First, if counties were able to anticipate the regulation and able to manipulate their 1974 TSPs levels to fall just below the cutoff, this bunching behavior would mean that counties just on either side of the cutoff value may not have the same potential outcomes. Another way this assumption could fail is if one or more of the other determinants of housing prices jumped discontinuously at the cutoff.

Figure 3 plots the results of estimating the nonparametric bivariate relationship between pollution changes and 1974 TSPs levels and housing price changes and 1974 TSPs levels using a bandwidth of 0.8 in the Stata `lowess` command. The LATE estimate here is the ratio between (i) the vertical distance between the two lines at the cutoff value in bottom panel, and (ii) the vertical distance between the two lines at the cutoff value in the top panel. This is similar to dividing the reduced form coefficient by the first-stage coefficient in a typical 2SLS regression.

Contrary to part 4, the top panel of Figure 3 shows that increases in the probability of regulation for counties near the TSPs cutoff are associated with increases in pollution from 1969-72 to 1977-80. Also contrary to part 4, the bottom panel of Figure 3 shows that increases in the probability of regulation for counties near the TSPs cutoff are associated with decreases in housing values from 1970 to 1980. These magnitudes are relatively small however, and might not be statistically significant if I were to estimate these relationships in a more formal way. However, when we divide the two vertical distances at the cutoff value as described above, we obtain a LATE estimate for the effect of pollution on housing values that has the same sign as the LATE estimate in part 4.

6.

Figure 4 plots (i) the nonparametric bivariate relation between the single-index measure of the housing price changes predicted to occur due to other variables changing and 1974

TSPs levels (red line) against (ii) the smoothed housing price changes from part 5 (blue line). If the continuity assumptions held, we would expect the index to be smooth around our cutoff value (i.e., not jump discontinuously) since it is constructed from our controls. Because the red line does jump discontinuously at the cutoff in a similar magnitude to jump in the blue line, this suggests that our continuity assumptions may not hold here.

7.

Figure 5 plots the results for this problem. These plots suggest different effects than the plots in Figure 3. That is, the plots in Figure 5 show that for counties near the cutoff, regulation is associated with a decrease in pollution values and increase in housing values. These effects match the signs of our estimates in part 4 and match the broader story that regulation decreased pollution and subsequently increased housing values for counties below the pollution cutoff.

8.

In general, 2SLS only recovers the LATE, or the ATE for the subpopulation whose treatment status is influenced by the instrumental variables (i.e., the compliers). This effect is identified under our usual IV assumptions of the exclusion restriction and a non-zero first stage. When one uses EPA regulation as an instrumental variable for air pollution, the LATE estimates reflect how housing values are impacted by changes in air pollution that occur because of the changes in EPA regulation.

9.

There are three key takeaways from the preceding analysis. First, the OLS estimates of the relationship between TSPs and housing values likely suffer from omitted variables bias and selection bias. These estimates say that increases in pollution increase housing values *ceteris paribus*, which is the opposite of what we would expect.

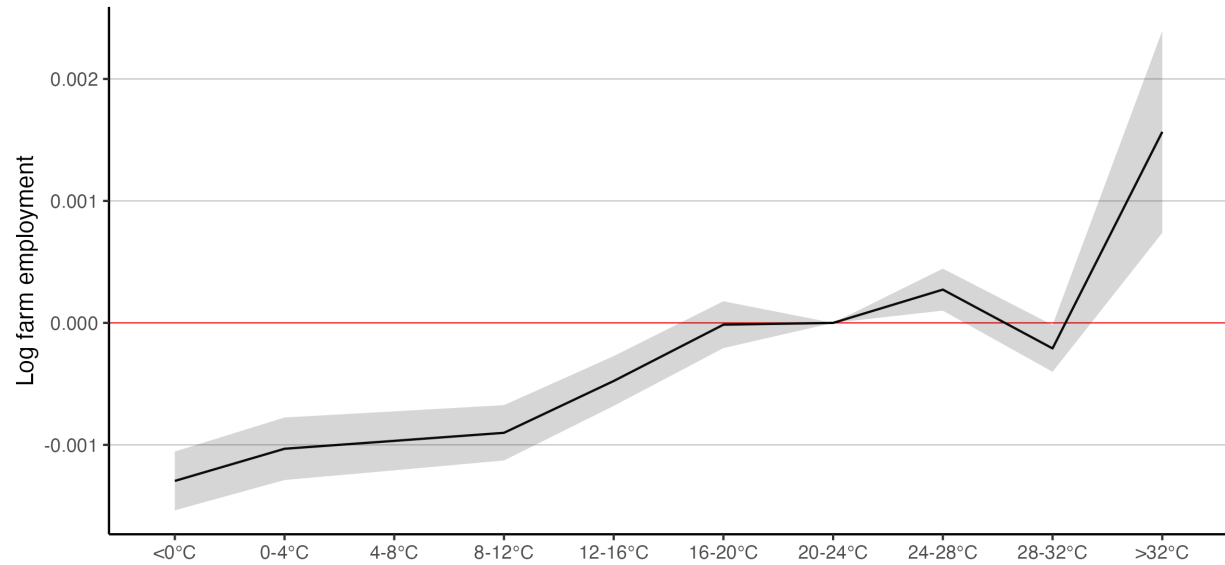
Second, the 2SLS estimates are consistent with our priors - regulation decreases TSPs and, through this channel, increases housing values *ceteris paribus*. The 2SLS estimates in parts 3 and 4 are also very similar in magnitude to one another despite using different instruments for changes in pollution. This may give us more confidence in the exclusion restriction holding, although we can never be certain. For instance, regulation may not affect housing values solely through the pollution channel if it also results in the economic displacement of firms or home buyers.

Lastly, the RD estimates for the local average treatment effect of changes in TSPs on changes in housing values are the same sign as our IV estimates. However, only the RD estimates in part 7 produces reduced form and first-stage estimates that have the same sign as our IV estimates. The estimates in part 7 are perhaps more plausible than those in part 5 since we are not relying on the continuity assumption holding across the cutoff value. Regulation is also perhaps more plausibly exogenous below the cutoff value since, conditional on mean TSPs in 1974, counties that were vs were not regulated only differed in whether their 2nd highest daily concentration above 260 units in 1974.

Because these are hedonic regressions, we can interpret them as saying that people positively value lower levels of TSPs insofar as those values are reflected in housing prices. In other words, people have a positive WTP to live in less areas with lower levels of TSPs. It is important to keep in mind, however, that without assuming homogeneous treatment effects, our IV and RD estimates reflect *local* average treatment effects, not average treatment effects. This means that the estimates are driven by compliers, or the population of counties for which regulation induces changes in pollution levels. To the extent that there is substantial heterogeneity in the population, these estimates may not be representative of WTP for air quality in other settings. Moreover, TSPs are not the only class of pollutant, and these estimates are therefore not reflective of overall WTP for reductions in air pollution.

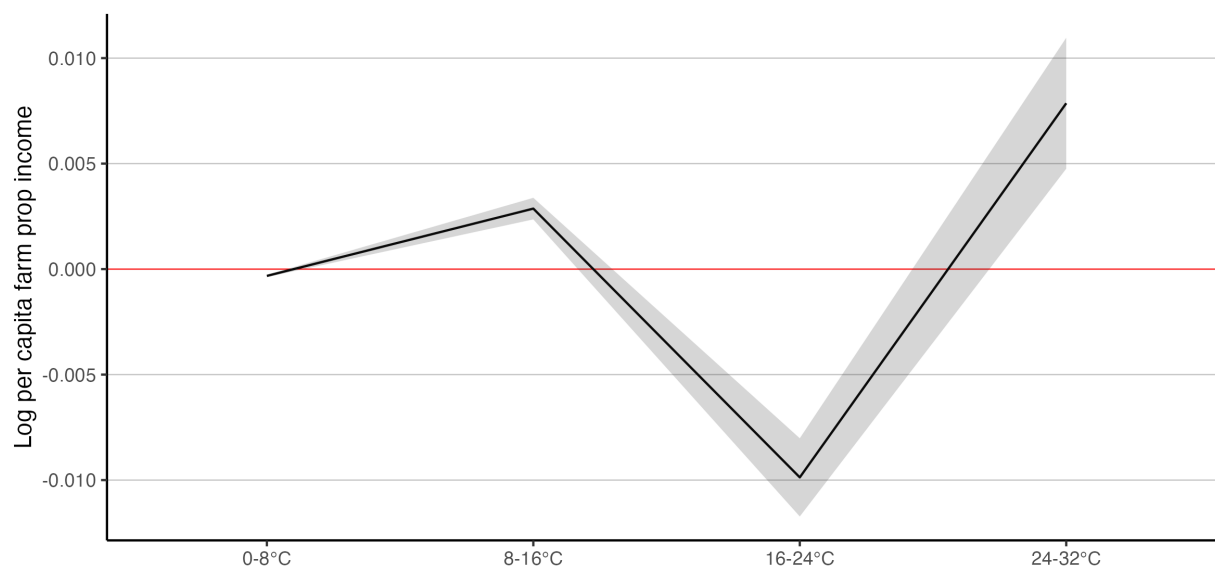
Figures

Figure 1: Exercise 1.2.1



The solid black line shows the point estimates of the regression specified in the problem. Gray shaded areas display 95% confidence intervals obtained using robust standard errors.

Figure 2: Exercise 1.2.2



The solid black line shows the point estimates of the regression specified in the problem. Gray shaded areas display 95% confidence intervals obtained using robust standard errors.

Figure 3: Exercise 2.5, TSPs changes (top) and housing price changes (bottom) versus 1974 TSPs levels

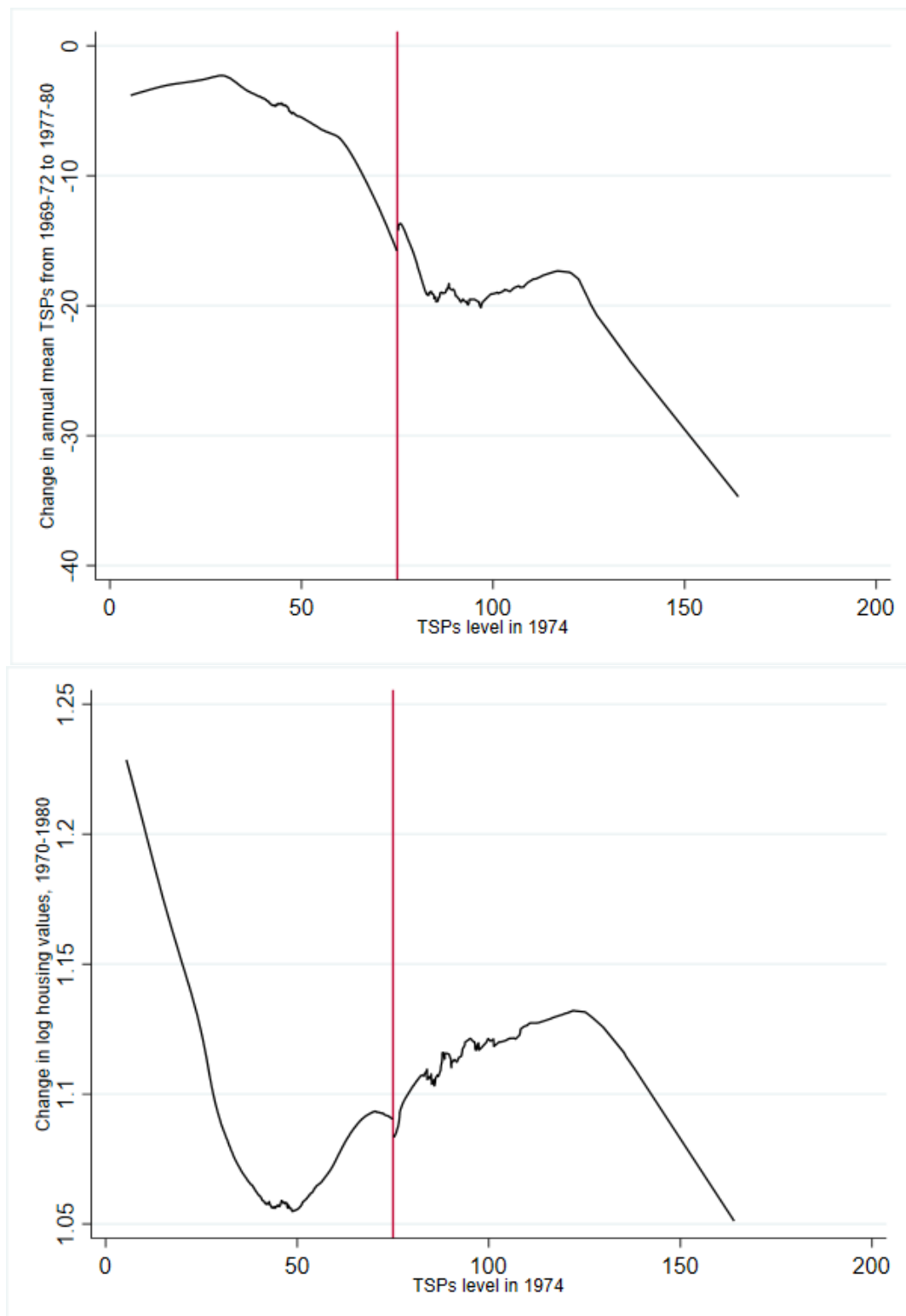
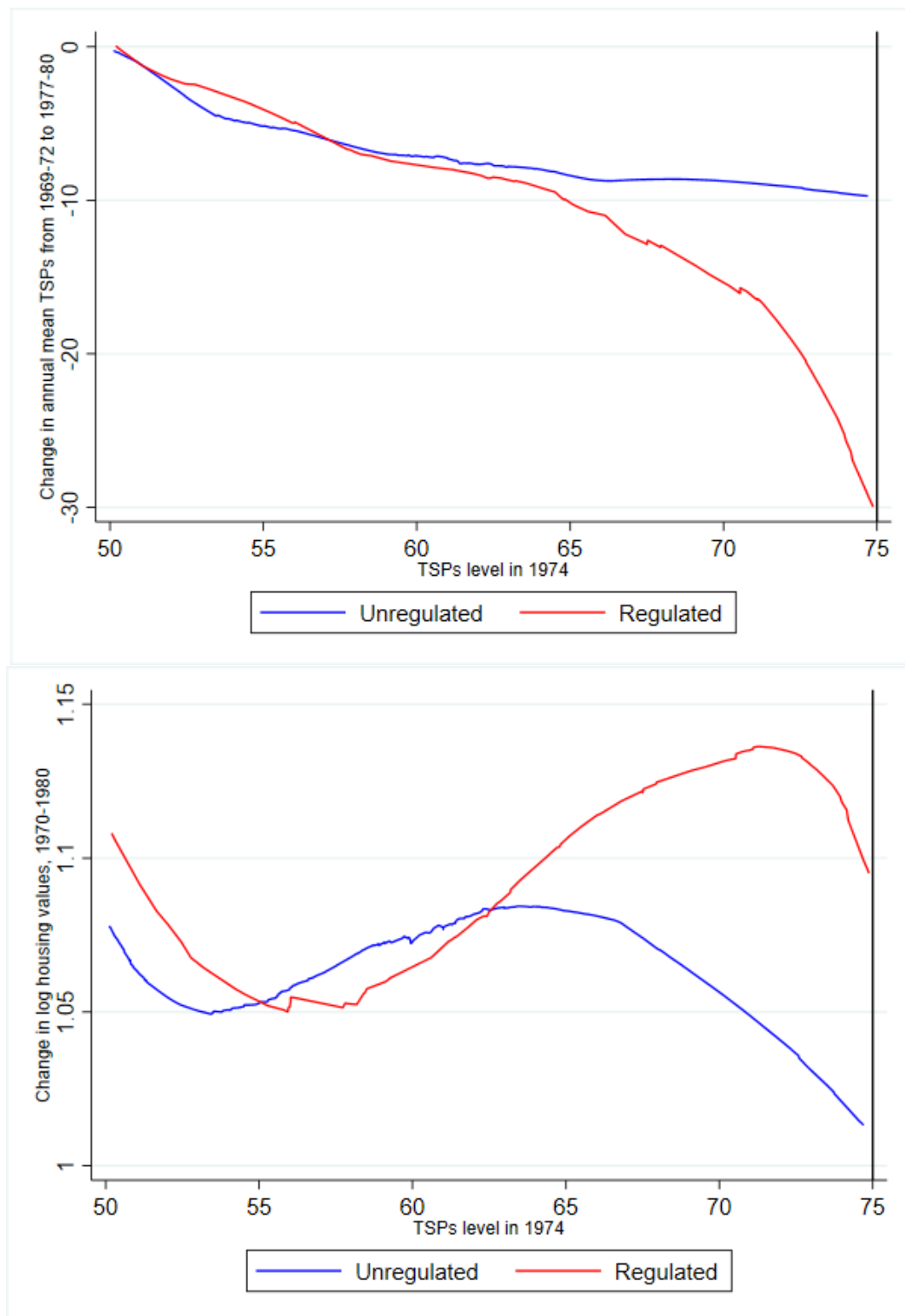


Figure 4: Exercise 2.6, TSPs changes and 1974 TSPs levels



Figure 5: Exercise 2.7, TSPs changes (top) and housing price changes (bottom) versus 1974 TSPs levels



Tables

Table 1: Exercise 1.2.3

	(1) log_emp_farm
tempB0	-0.001300 (0.000123)
temp0to4	-0.000998 (0.000131)
temp4to8	-0.001020 (0.000124)
temp8to12	-0.000946 (0.000116)
temp12to16	-0.000493 (0.000105)
temp16to20	-0.000016 (0.000098)
temp20to24	
temp24to28	0.000272 (0.000088)
temp28to32	-0.000251 (0.000099)
tempA32	-0.002022 (0.008176)
tempB0 \times tempA32	-0.000184 (0.000047)
temp0to4 \times tempA32	-0.000117 (0.000079)
temp4to8 \times tempA32	0.000201 (0.000050)
temp8to12 \times tempA32	0.000034 (0.000037)
temp12to16 \times tempA32	0.000052 (0.000025)
temp16to20 \times tempA32	-0.000001 (0.000035)
temp24to28 \times tempA32	0.000050 (0.000040)
temp28to32 \times tempA32	-0.000031 (0.000028)
tempA32 \times tempA32	-0.000046 (0.000033)
constant	6.835019 (0.024687)
N	126619
R^2	0.948

Robust standard errors are reported in parentheses.

Table 2: Exercise 2.1

	(1) dlhouse	(2) dlhouse	(3) dgtsp	(4) dlhouse
dgtsp	0.00264 (0.00084)	0.00038 (0.00037)		
dincome		0.00010 (0.00002)	0.00800 (0.00191)	0.00011 (0.00004)
dunemp		-4.16578 (0.73011)	-32.43427 (67.96615)	-3.96449 (1.15538)
dmnfcg		-0.83358 (0.28736)	109.81310 (30.95433)	0.33171 (0.26321)
constant	1.07594 (0.01468)	1.09699 (0.07045)	-17.89767 (3.12364)	1.03802 (0.06511)
Controls	No	Yes	No	No
Weights	Yes	Yes	Yes	Yes
N	967	967	967	967
R^2	0.064	0.600	0.140	0.426

Robust standard errors are reported in parentheses. Controls refers to all “other relevant variables” listed in the data notes aside from the three economic shocks: change in income per-capita (dincome), change in unemployment rate(dunemp), and change in percent manufacturing employment (dmnfcg).

Table 3: Exercise 2.2

	(1) tsp7576
dincome	-0.00006 (0.00009)
dunemp	0.61663 (2.92262)
dmnfcg	-0.51171 (0.88861)
constant	0.49508 (0.14339)
Controls	No
Weights	Yes
N	967
R^2	0.010

Robust standard errors are reported in parentheses.

Table 4: Exercise 2.3

	(1) dgtsp	(2) dgtsp	(3) dlhouse	(4) dlhouse	(5) dlhouse	(6) dlhouse
tsp7576	-9.84390 (2.19659)	-8.05175 (1.78985)	0.04445 (0.03707)	0.03966 (0.01754)		
dgtsp					-0.00452 (0.00407)	-0.00493 (0.00247)
Controls	No	Yes	No	Yes	No	Yes
Weights	Yes	Yes	Yes	Yes	Yes	Yes
N	967	967	967	967	967	967
R^2	0.067	0.223	0.013	0.607	-0.406	0.390

The dependent variable is shown at the top of each column. Columns 1 and 2 correspond to the first-stage regressions, Columns 3 and 4 correspond to the reduced form regressions, and Columns 5 and 6 correspond to the 2SLS regressions. Robust standard errors reported in parentheses.

Table 5: Exercise 2.4

	(1) dgtsp	(2) dgtsp	(3) dlhouse	(4) dlhouse	(5) dlhouse	(6) dlhouse
mtspgm74	-0.27009 (0.05738)	-0.24288 (0.05128)	0.00124 (0.00089)	0.00124 (0.00039)		
dgtsp					-0.00459 (0.00356)	-0.00512 (0.00196)
Controls	No	Yes	No	Yes	No	Yes
Weights	Yes	Yes	Yes	Yes	Yes	Yes
N	967	967	967	967	967	967
R^2	0.082	0.231	0.016	0.610	-0.415	0.374

The dependent variable is shown at the top of each column. Columns 1 and 2 correspond to the first-stage regressions, Columns 3 and 4 correspond to the reduced form regressions, and Columns 5 and 6 correspond to the 2SLS regressions. Robust standard errors reported in parentheses.