Internet Appendix for:

Eye in the sky: private satellites and government macro data

Abhiroop Mukherjee, George Panayotov, Janghoon Shon

April 2020

IA-1. Measuring cloudiness in the U.S.

We collect high-frequency sky coverage data from NOAA (`https://www.ncei.noaa.gov/data/global-hourly/`), for several locations. This data is part of the Cloud and Solar Data module and represents the fraction of the sky covered by various types of clouds. We normalize it to take values between zero (completely clear) and one (completely cloudy).

The cloud coverage data we use are all recorded at airports, where generally such data is of higher quality. In particular, for Cushing we use data from the Cushing Municipal Airport (closest, but only available for 2010-2018) and Stillwater Regional airport (2007-2009); for Houston - Houston William P. Hobby Airport (2007-2018); for Patoka - Salem Leckrone Airport (2010-2018) and Mount Vernon Airport (2007-2009); for Midland - Midland International Airport (2007-2018); for Houma - Houma Terrebonne Airport (2007-2018); for Corpus Christi - Corpus Christi International Airport (2007-2018); for Beaumont - Port Arthur Regional Airport (2007-2018); for Wichita Falls - Wichita Falls Municipal Airport (2007-2018); for Wink - Winkler County Airport (2007-2018), and for Baton Rouge - Baton Rouge Ryan Airport (2007-2018). Using the same airport throughout for each location does not change our results.

We note that the Earth-observation satellites, which are the source of imagery used to estimate oil inventories, typically have low Earth orbits and hence revolve around the Earth multiple times a day. In many cases, their orbits pass over given points at approximately the same time each day (e.g., `https://earthobservatory.nasa.gov/features/OrbitsCatalog`). However, we have no information about the exact times when a satellite passes over locations of interest to us. In addition, multiple satellites may be used to monitor each oil storage location, each passing over it at different times. Therefore, we obtain a daily cloudiness measure for each location by averaging sky coverage values over the daylight period. For our main result we use 7:00 to 18:00 as the daylight period, and show robustness in Table IA-3. Next, we average these values across the selected locations to obtain an aggregate cloudiness measure for a given day. Our construction of weekly cloudiness based on these daily measures is described in Section 4.2 in the paper.

IA-2. Robustness of the U.S. crude oil results

Table IA-3 establishes the robustness of our findings for the U.S. oil market. The top panel in the table refers to the cutoff in the weekly cloudiness measure that we use to separate clear from cloudy weeks. Instead of the baseline cutoff at the 75th percentile, this panel shows results obtained using the 65th and 85th percentiles. The regression coefficient estimates and their significance do not differ materially from those in Table 2 in the paper. The difference between the respective β_{clear} and β_{cloudy} estimates remains significant. The first panel also demonstrates robustness from a different angle and considers different "expected" oil inventory changes that enter the calculation of the unexpected oil inventory changes ΔOil_Inv_t in equation (1). While in Table 2 we use the average change in inventories in the preceding four weeks, here we show results using the average change over the preceding one or 13 weeks (one quarter) to represent the inventory expectations. The impact of such changes is minimal, whereby both the coefficient estimates and their difference remain close to their baseline values.

The second panel of Table IA-3 shows that robustness is preserved if (i) one more year is added to the baseline and pre-period (i.e. it becomes 2013-2018 and 2007-2012), or (ii) we additionally control for $Clear_t$ in the regression specification. The third panel of the table shows results when cloudiness is calculated using cloud data from 9:00 to 15:00, or from 10:00 to 14:00 (Daylight 2 and 3, respectively; elsewhere in the paper we use 7:00 to 18:00). Finally, the column denoted "incl. Wed" ("excl. Thu") shows results where we measure cloudiness starting on the Wednesday (Friday) preceding the announcement (instead of the Thursday as in our main results).

Overall, Table IA-3 demonstrates that our baseline assumptions are not crucial for our main results, which remain intact even under significant changes in some of these assumptions.

IA-3. Measuring cloudiness in China

For China, we obtain cloudiness data from the same database as we use for the US (NOAA's Surface Data Hourly Global), which is a worldwide collection of surface weather observations. The stations from which we source cloudiness data for the few Chinese industrial hubs are selected based on their proximity to the hubs and the completeness of their sky-coverage data.

We use data from several airports: for Nanjing - Nanjing Lukou International Airport, for Guangzhou - Guangzhou Baiyun International Airport, for Qingdao - Qingdao Liuting International Airport, and for Hangzhou - Hangzhou Xiaoshan International Airport. In robustness checks, we also use Shanghai Hongqiao International Airport for Shanghai and Shenzhen Baoan International Airport for Shenzhen. In addition, we use data from stations at Liyang, Longkou - for Yantai, Quxian - for Jinhua, and Gaoyao - for Zhaoqing.

IA-4. Robustness of the Chinese PMI results

Similar to Table IA-3, Table IA-8 establishes the robustness of our findings for the Chinese stock market and PMI. The top panel refers to the cutoff in the monthly cloudiness measure that separates clear from cloudy months and shows results obtained with the 65th and 85th percentiles, instead of the baseline cutoff at the 75th percentile. The regression results are very similar to those in Table 9 in the paper, with significant β_{cloudy} estimates. The top panel also confirms the robustness of the results when the expected PMI is calculated as a different moving average. The difference between the two slope estimates remains insignificant, except in one case.

The bottom panel of Table IA-8 shows that robustness is preserved if we subtract one year from the baseline and pre-periods (i.e., if we use 2015-2018 and 2009-2012 instead). Furthermore, the results remain intact if we include the two cities that host the main Chinese stock exchanges (Shanghai and Shenzhen) in the calculation of the cloudiness measure, which now averages across ten, and not eight cities. Such a change is meaningful, as both cities are situated in, or immediately next to, one of the four provinces where Chinese manufacturing is concentrated.

In sum, Table IA-8 demonstrates that the results obtained with our baseline assumptions are not particularly sensitive to various modifications of these assumptions.

Table IA-1

Impact of cloudiness on the U.S. oil market and other macro variables

This table shows results from testing the relation between local cloudiness over our selected oil storage hubs and several oil market and macroeconomic variables. We regress these variables on a constant and a dummy variable, that equals one in cloudy weeks, and zero otherwise. Bootstrap p-values are in parentheses. The oil returns are as in Table 2 in the paper, and oil inventory changes (reported in percent) are calculated from the EIA's weekly oil inventory announcements. The remaining variables are the weekly percentage changes in the S&P500, the Baltic Dry Index (BDI), and the Reuters CRB Continuous Commodity Indexes (Energy and Industrials), all scaled to unit standard deviation. If a Friday price is missing (e.g., a holiday), the Thursday price is used for calculating the weekly returns; if both the Thursday and Friday prices are missing, the respective week is dropped. We use the combined baseline and pre-period sample.

	Oil returns				**Oil inventory**
	10:30-11:00	10:00-11:00	09:45-11:15	09:30-11:30	EIA
Cloudy dummy	-0.09	-0.12	-0.16	-0.17	0.08
	(0.39)	(0.35)	(0.22)	(0.19)	(0.52)
R^2	0.00	0.00	0.00	0.00	0.00
Observations	522	522	522	522	522

	Other macroeconomic variables			
	S&P500	BDI	CRB (Energy)	CRB (Industrials)
Cloudy dummy	-0.12	-0.74	-0.31	0.13
	(0.62)	(0.40)	(0.43)	(0.69)
R^2	0.00	0.00	0.00	0.00
Observations	522	502	521	521

Table IA-2
Robustness: U.S. oil market I

This table replicates the results in Table 2 in the paper, with two differences. In the top panel, now both the left-hand and right-hand variables in the regressions are winsorized at the 1st and 99th percentiles. The bottom panel shows p-values obtained with robust standard errors, instead of bootstrap p-values.

Winsorized

	10:30-11:00	10:00-11:00	09:45-11:15	09:30-11:30
β_{clear}	-0.05	-0.10	-0.07	-0.08
	(0.50)	(0.18)	(0.36)	(0.32)
β_{cloudy}	-0.53***	-0.56***	-0.55***	-0.60***
	(0.00)	(0.00)	(0.00)	(0.00)
Difference	-0.48***	-0.46***	-0.48***	-0.52***
	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.05	0.06	0.04	0.05
Observations	261	261	261	261

Robust standard errors

	10:30-11:00	10:00-11:00	09:45-11:15	09:30-11:30
β_{clear}	-0.05	-0.10	-0.06	-0.07
	(0.51)	(0.19)	(0.48)	(0.45)
β_{cloudy}	-0.51***	-0.55***	-0.52***	-0.55***
	(0.00)	(0.00)	(0.00)	(0.00)
Difference	-0.46***	-0.45***	-0.46***	-0.48***
	(0.00)	(0.01)	(0.01)	(0.01)
R^2	0.05	0.06	0.04	0.04
Observations	261	261	261	261

Table IA-3
Robustness: U.S. oil market II

In the format of Table 2 in the paper, we show results using oil returns from 9:30 to 11:30 a.m. on EIA announcement days. The results in the first panel are obtained with two different percentile cutoffs (in weekly cloudiness) that separate clear from cloudy weeks, and with expected oil inventory change calculated as the average in the preceding one or 13 weeks. The second panel shows results for extended baseline and pre-periods, and from regressions that also include $Clear_t$ as a separate regressor. For the third panel, cloudiness is estimated over shorter daylight periods (09:00-15:00 and 10:00-14:00), and weekly cloudiness is calculated starting on the Wednesday or Friday preceding an EIA announcement, instead of the Thursday.

	65th pctile	85th pctile	MA(1)	MA(13)
β_{clear}	-0.07	-0.13	-0.12	-0.04
	(0.45)	(0.12)	(0.16)	(0.70)
β_{cloudy}	-0.46***	-0.65***	-0.50***	-0.38**
	(0.00)	(0.00)	(0.01)	(0.01)
Difference	-0.40***	-0.52**	-0.38**	-0.34*
	(0.01)	(0.04)	(0.05)	(0.05)
R^2	0.04	0.03	0.04	0.02
Observations	261	261	261	261
			$Clear_t$ included	
	2013-2018	2007-2012	2014-2018	2007-2011
β_{clear}	-0.04	-0.46***	-0.07	-0.51***
	(0.61)	(0.00)	(0.40)	(0.00)
β_{cloudy}	-0.44***	-0.48***	-0.55***	-0.47***
	(0.00)	(0.00)	(0.00)	(0.01)
Difference	-0.40***	-0.02	-0.48***	0.03
	(0.00)	(0.96)	(0.01)	(0.86)
R^2	0.03	0.12	0.04	0.13
Observations	313	313	261	261
	Daylight 2	Daylight 3	incl. Wed	excl. Thu
β_{clear}	-0.06	-0.06	-0.09	-0.12
	(0.54)	(0.52)	(0.34)	(0.19)
β_{cloudy}	-0.57***	-0.59***	-0.50***	-0.47***
	(0.00)	(0.00)	(0.00)	(0.00)
Difference	-0.51***	-0.53***	-0.41**	-0.35*
	(0.00)	(0.00)	(0.03)	(0.07)
R^2	0.04	0.05	0.04	0.03
Observations	261	261	261	261

Table IA-4

An alternative measure of expected U.S. oil inventory changes

In the format of Table 2 in the paper, this table shows the results from similar regressions, again for the EIA's announcement days, but now ΔOil_Inv_t is calculated using the expected change derived from the oil inventory change announced by the American Petroleum Institute (API). API data is only available for the baseline period (at Datastream).

	(2014-2018)			
	10:30-11:00	10:00-11:00	09:45-11:15	09:30-11:30
β_{clear}	-0.04	-0.08	-0.07	-0.06
	(0.64)	(0.32)	(0.44)	(0.60)
β_{cloudy}	-0.35**	-0.39***	-0.44***	-0.47***
	(0.01)	(0.01)	(0.00)	(0.00)
Difference	-0.31**	-0.31*	-0.37**	-0.41**
	(0.04)	(0.08)	(0.04)	(0.03)
R^2	0.03	0.04	0.04	0.04
Observations	261	261	261	261

Table IA-5

Subsamples of the baseline period: U.S. oil market

This table replicates the results in Table 2 in the paper, with the only difference that now we use separately data from the first two and last three years of the baseline period.

		(2014-2015)		
	10:30-11:00	10:00-11:00	09:45-11:15	09:30-11:30
β_{clear}	-0.14	-0.20**	-0.20*	-0.27**
	(0.13)	(0.04)	(0.07)	(0.02)
β_{cloudy}	-0.57***	-0.62***	-0.72***	-0.83***
	(0.01)	(0.00)	(0.00)	(0.00)
Difference	-0.43*	-0.42*	-0.51**	-0.56*
	(0.09)	(0.09)	(0.05)	(0.06)
R^2	0.09	0.11	0.09	0.10
Observations	105	105	105	105
		(2016-2018)		
	10:30-11:00	10:00-11:00	09:45-11:15	09:30-11:30
β_{clear}	0.02	-0.02	0.06	0.08
	(0.87)	(0.90)	(0.59)	(0.55)
β_{cloudy}	-0.47**	-0.50**	-0.42*	-0.40**
	(0.02)	(0.02)	(0.07)	(0.03)
Difference	-0.48**	-0.48**	-0.47*	-0.48**
	(0.03)	(0.05)	(0.06)	(0.04)
R^2	0.04	0.05	0.03	0.03
Observations	156	156	156	156

Table IA-6

Impact of cloudiness on the Chinese stock market and macro variables

This table show results from regressing Chinese stock market returns and other macro variables on a constant and a dummy variable, that equals one in cloudy months, and zero otherwise. Bootstrap p-values are in parentheses. The CSI300 index returns are as in Table 9, and PMI denotes the Chinese manufacturing PMI. The remaining variables are the monthly percentage changes in the Shanghai Stock Exchange Composite Index (SSE), the Baltic Dry Index (BDI), and the Reuters CRB Continuous Commodity Indexes (Energy and Industrials), all scaled to unit standard deviation. Shown are also R^2's and the number of (monthly) observations for each variable. We use the combined baseline and pre-period sample.

	CSI300 stock index returns				**PMI**
	close-10:00	close-10:30	close-11:00	close-11:30	
Cloudy dummy	-0.04	0.02	0.06	0.03	0.01
	(0.84)	(0.99)	(0.72)	(0.82)	(0.74)
R^2	0.00	0.00	0.00	0.00	0.00
Observations	120	120	120	120	120

	Other macroeconomic variables			
	Shanghai Composite	BDI	CRB (Energy)	CRB (Industrials)
Cloudy dummy	0.10	-0.15	0.24	0.15
	(0.57)	(0.47)	(0.21)	(0.47)
R^2	0.00	0.00	0.01	0.00
Observations	120	120	120	120

Table IA-7
Robustness: Chinese stock index returns I

This table replicates the results in Table 9 in the paper, with two differences. In the top panel, now both the left-hand and right-hand variables in the regressions are winsorized at the 1st and 99th percentiles. The bottom panel shows p-values obtained with robust standard errors, instead of bootstrap p-values.

	Winsorized			
	close-10:00	close-10:30	close-11:00	close-11:30
β_{clear}	0.22	0.20	0.25*	0.08
	(0.16)	(0.23)	(0.09)	(0.70)
β_{cloudy}	0.50***	0.53***	0.48**	0.57**
	(0.00)	(0.00)	(0.04)	(0.03)
Difference	0.28	0.33	0.23	0.49*
	(0.19)	(0.18)	(0.38)	(0.10)
R^2	0.09	0.06	0.07	0.04
Observations	60	60	60	60
	Robust standard errors			
	close-10:00	close-10:30	close-11:00	close-11:30
β_{clear}	0.21	0.20	0.25	0.08
	(0.13)	(0.20)	(0.11)	(0.62)
β_{cloudy}	0.39***	0.40***	0.38***	0.44***
	(0.00)	(0.00)	(0.01)	(0.01)
Difference	0.18	0.21	0.13	0.36
	(0.55)	(0.57)	(0.72)	(0.37)
R^2	0.07	0.05	0.06	0.03
Observations	60	60	60	60

Table IA-8

Robustness: Chinese stock index returns II

In the format of Table 9 in the paper, this table shows results using CSI300 returns from the close on the previous day to 11:30 a.m. on each PMI announcement day. In the top panel, the first two columns are obtained with two different cutoffs (in monthly cloudiness) that separate clear from cloudy months. For the last two columns, the expected PMI is calculated as the average PMI in the preceding three and nine months. In the bottom panel, the first two columns show results for the shorter sample periods 2015-2018 and 2009-2012. The last two columns show results when the cloudiness measure is calculated including Shanghai (SH) and Shenzhen (SZ).

	65th pctile	85th pctile	MA(3)	MA(9)
β_{clear}	0.04	0.10	0.13	0.10
	(0.81)	(0.54)	(0.52)	(0.49)
β_{cloudy}	0.53***	0.50**	0.48**	0.40**
	(0.01)	(0.04)	(0.03)	(0.03)
Difference	0.48*	0.41	0.34	0.30
	(0.06)	(0.14)	(0.19)	(0.19)
R^2	0.04	0.03	0.04	0.03
Observations	60	60	60	60

			SH and SZ included	
	2015-2018	2009-2012	2014-2018	2009-2013
β_{clear}	0.23	0.45	0.09	0.41
	(0.22)	(0.12)	(0.61)	(0.15)
β_{cloudy}	0.51**	0.38*	0.44**	0.36*
	(0.04)	(0.10)	(0.05)	(0.10)
Difference	0.28	-0.07	0.35	-0.05
	(0.27)	(0.84)	(0.18)	(0.89)
R^2	0.06	0.09	0.03	0.08
Observations	48	48	60	60

Table IA-9

Subsamples of the baseline period: Chinese stock index

This table replicates the results in Table 9 in the paper, with the only difference that now we use separately data from the first two and last three years of the baseline period.

	(2014-2015)			
	close-10:00	close-10:30	close-11:00	close-11:30
β_{clear}	0.15	0.11	0.18	-0.12
	(0.50)	(0.64)	(0.46)	(0.59)
β_{cloudy}	0.33*	0.24	0.24	0.26
	(0.10)	(0.24)	(0.31)	(0.43)
Difference	0.18	0.13	0.06	0.38
	(0.59)	(0.67)	(0.73)	(0.32)
R^2	0.02	0.01	0.02	0.01
Observations	24	24	24	24

	(2016-2018)			
	close-10:00	close-10:30	close-11:00	close-11:30
β_{clear}	0.25	0.25	0.32	0.25
	(0.16)	(0.19)	(0.11)	(0.27)
β_{cloudy}	0.42**	0.48**	0.49**	0.60**
	(0.04)	(0.03)	(0.05)	(0.03)
Difference	0.18	0.23	0.17	0.34
	(0.46)	(0.38)	(0.53)	(0.27)
R^2	0.15	0.14	0.16	0.12
Observations	36	36	36	36

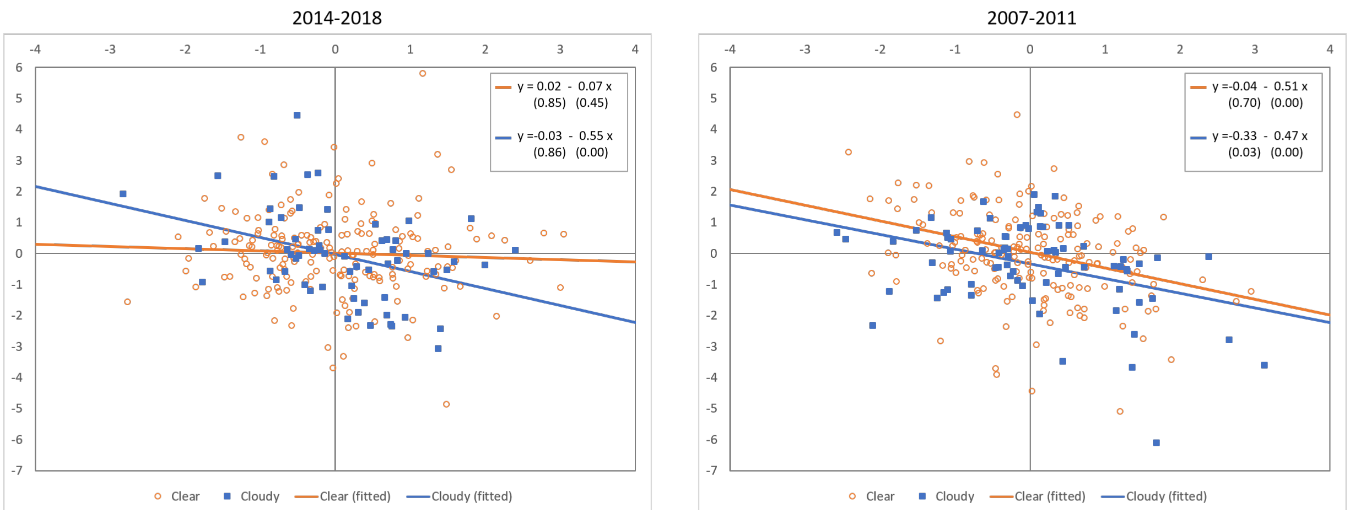


Figure IA-1: **Scatterplots – U.S. oil market**

The left panel of the figure shows scatter plots and fitted regression lines for clear and cloudy weeks (with orange circles and blue squares, and an orange and blue line, resp.). The horizontal axis shows the (unexpected) change in U.S. oil inventories (ex-SPR), scaled to unit standard deviation, over the baseline period (2014-2018). The vertical axis shows the front-month oil (WTI) futures returns from 9:30 a.m. to 11:30 a.m. on EIA announcement days over the same period. These returns are shown in percent. The estimated regression equations are also displayed. The right panel of the figure reproduces the same plot, but for the pre-period (2007-2011).

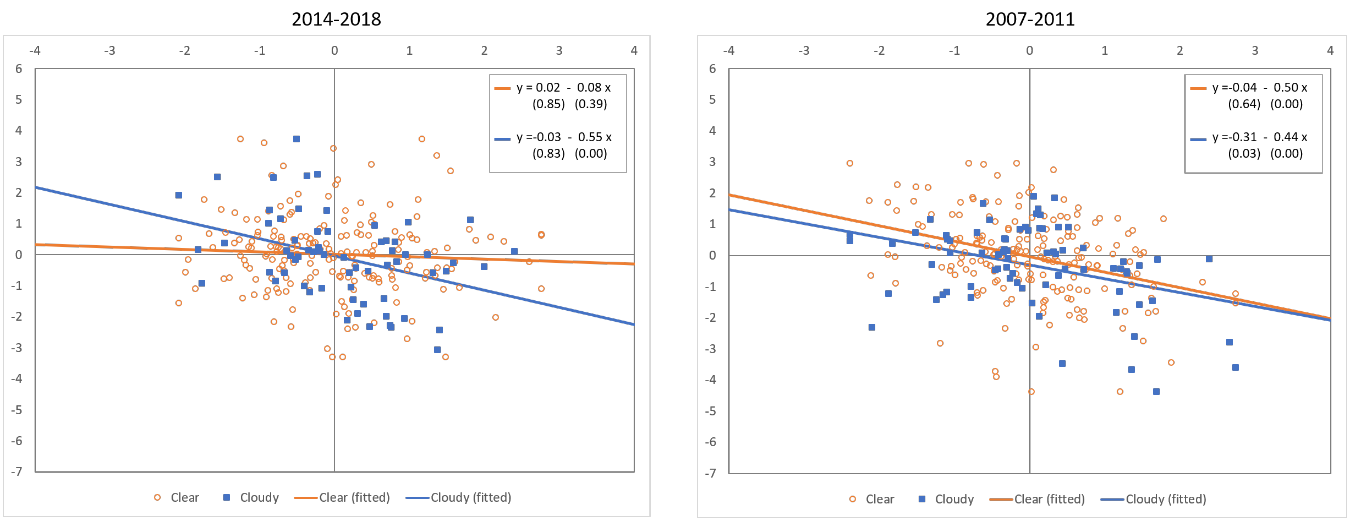
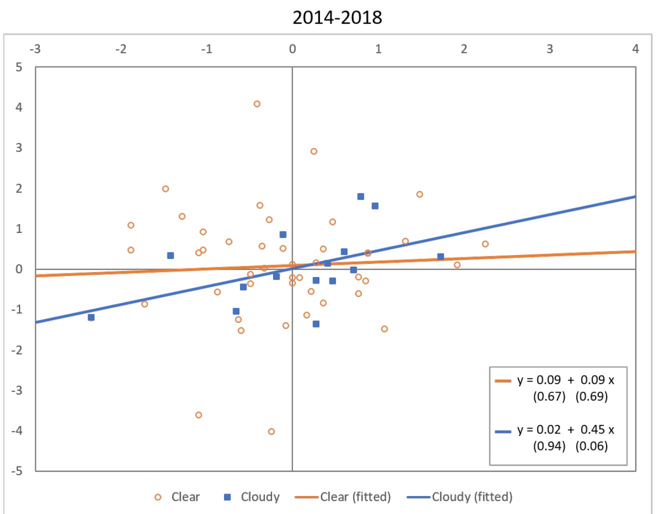


Figure IA-2: **Scatterplots – U.S. oil market (winsorized)**

This figure replicates the scatter plots and regression lines from Figure IA-1, with the only difference that now both the left-hand and right-hand variables in the regressions are winsorized at the 1st and 99th percentiles.



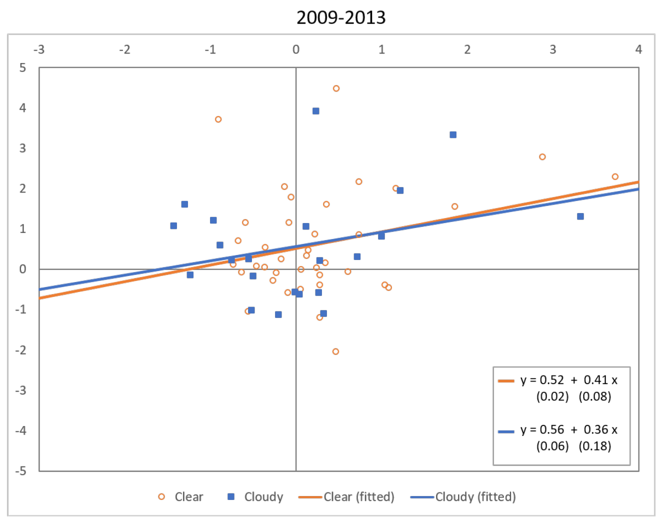


Figure IA-3: **Scatterplots – Chinese stock market**

This figure displays scatter plots and regression lines similar to those in Figure IA-1, but now the horizontal axis in each plot shows the (unexpected) changes in the Chinese manufacturing PMI, scaled to unit standard deviation, and the vertical axis shows the CSI300 returns (in percent) from the close on the day preceding a PMI announcement to 11:30 a.m. on the morning following such a PMI announcement. The baseline period here is 2014-2018 (left panel) and the pre-period is 2009-2013 (right panel).

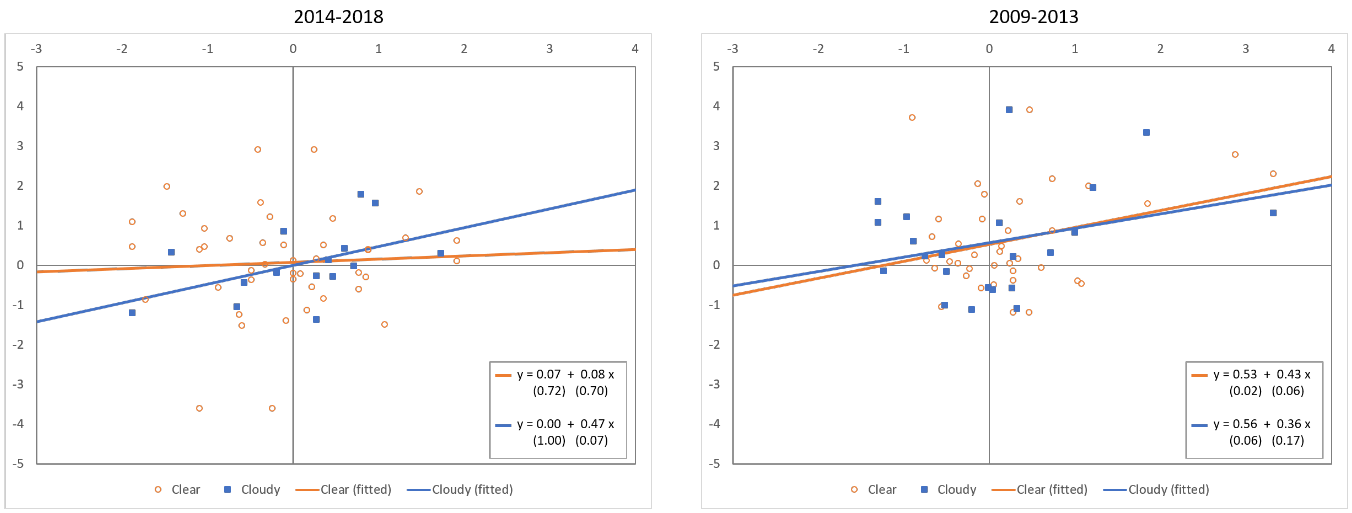


Figure IA-4: **Scatterplots – Chinese stock market (winsorized)**

This figure replicates the scatter plots and regression lines from Figure IA-3, with the only difference that now both the left-hand and right-hand variables in the regressions are winsorized at the 1st and 99th percentiles.