Temperature and Temperament: Appendix

Patrick Baylis

December, 2015*

[WORK IN PROGRESS]

This appendix contains supplemental material to **Baylis2015a**, organized as follows:

*The paper is available at: http://are.berkeley.edu/candidate/patrick-baylis.

1 Theory

1.1 Text analysis and measurement error

Given direct observations of human mood, a researcher could plausibly estimate the causal effect of temperature on mood with the following empirical specification:

$$E_{it} = f(T_{it}) + \phi_i + \phi_t + \epsilon_{it} \tag{1}$$

$f(T_{it})$ is some flexible function of temperature, while the fixed effects ϕ_i and ϕ_t account for unobserved, correlated variation in emotion and weather between units and time periods, respectively. The standard identifying assumption is that $E[T_{it}\epsilon_{it} = 0|\phi_i, \phi_t]$, i.e. that observed variation in temperature is as good as random after conditioning on the fixed effects.

However, to date there is no technology available to measure mood directly[1], and so E_{it} is unavailable to the researcher. The next best substitute, at least at large scale, is the rapidly developing field of sentiment analysis.

Sentiment analysis, or opinion mining, is a natural language processing technique designed to elicit subjective opinions from written or spoken source material. The technique was originally developed by computational linguists and is now widely used by both academic researchers and practitioners in the private sector (Pang and Lee 2008). Researchers have also used sentiment analysis to extract subjective opinions from Twitter status updates (Pak and Paroubek 2010). Work in sentiment analysis broadly falls into one of two cate-

[1]Edgeworth (1881) actually proposed a "hedonimeter", a device to directly assess an individual's present level of utility. It has not yet been invented.

gories: using estimated sentiment as a right-hand side variable to improve predictive models of outcomes of interest (Bollen, Mao, and Zeng 2011; Gerber 2014), and using sentiment as a left-hand side variable to understand the impact of a set of covariates on human mood (Dodds and Danforth 2010; Mitchell et al. 2013).

In the latter case, understanding of the potential bias introduced by the sentiment analysis process guides the interpretation of the estimated parameters. While sentiment analysis practitioners demonstrate an interest in minimizing this bias, no work has yet formalized the sentiment analysis process to better understand its properties in a setting where causal parameters are of scientific interest. I fill this gap by linking a simple model of sentiment analysis to the well-understood econometric theories of measurement error and omitted variable bias.

To model sentiment analysis, I begin by adapting the psycholinguistic theory of language production to the tweet generation process, which describes the production of spoken or written language as a series of stages whereby a intended message is formed in the brain, encoded into words, and expressed (either written or spoken)[2].

Consider a user who has already decided to tweet and chosen her intended message. In the lexical choice stage of language production, she then selects m, a message of 140 characters or less. Following psycholinguistic theory, the lexical choice function f is a function of the idea the user wishes to express I, and the user's present emotional state E, s.t. $m = f(I, E)$. For expositional purposes, let $I \in \mathbb{R}$, s.t. the set of possible ideas is uncountably infinite. Also let $E \in$ as a real number between 0 and 1. m is constrained to be finite[3], since there are a large but finite set of possible message.

A natural (to economists) model of f builds on the standard model of utility maximization, s.t. users choose $m = \arg\max_m U(m, E(T, \cdot), x(T, \cdot))$. A sufficient but not necessary "honesty" condition ensures that utility declines in the distance between expressed emotion

[2]In fact, there are many different theories of language production. I focus on the canonical serial models described by Garrett (1976) and Clark (1975).

[3]Empirically, practitioners of sentiment analysis usually collapse m to a count of its constituent words, but this distinction is unimportant for the model.

m and internal emotion E, i.e. $\frac{\partial U}{\partial |m-E|} < 0$. A less stringent condition suggests that the results of the model hold so long as individuals generally select words that reflect internal emotions in a similar way. In other words, if everyone lies in the same manner about their internal emotion, the effect is the same. The honesty condition is less general but more plausible.

Since f maps from an infinite domain to a finite range, f^{-1} does not exist. This implication of the model reflects reality: a given phrase may map to more than one message-emotional state combination.

Still, so long as $\frac{\partial m}{\partial E} \neq 0$, i.e. affective state affects word choice, then observing m supplies information on E. Suppose that a researcher observes both m and E for some subset of her data, perhaps by polling her sample on the sentiment they associate with different words m. By running the regression $E_{it} = \sum_m^M \beta_m \mathbb{1}[m_i = m] + \mu_i$, where the first set of regressors are a set of dummy variables for every possible m, she estimates β_m, the average emotional state for each expression m.

Now suppose that the researcher observes only m for the remainder of the data. She can use the β_m estimates to predict $\hat{E}_i = \sum_m^m \hat{\beta_m} \mathbb{1}[m_i = m]$. Note that $E_{it} - \hat{E}_{it} = \mu_{it}$. Returning to equation 1, we substitute $\hat{E}_{it}$ for E_{it}. Using that $\varepsilon_i = \mu_i + \epsilon_i$, our model is as follows:

$$\hat{E}_{it} = f(T_{it}) + \phi_i + \phi_t + \overbrace{\mu_{it} + \epsilon_{it}}^{\varepsilon_{it}} \tag{2}$$

ϵ_{it} is our original error term, which in the current model is uncorrelated with the covariates after controlling for the appropriate fixed effects per the literature. μ_{it} is the measurement error term, and ε_{it} is the composition of the two. The addition of μ_{it} captures the additional variance added by sentiment analysis. Following Wooldridge 2002, we can characterize this as measurement error in our left-hand side variable. The crucial question is whether or not μ_{it} correlates with the right-hand side variable.

1. If μ_{it} is uncorrelated with T_{it} after conditioning on the fixed effects, then the estimate

is unbiased but less efficient.

2. If μ_{it} is correlated with T_{it} after conditioning on the fixed effect, then the estimate has a classical omitted variables problem.

Under the assumption of uncorrelated measurement error, mood as estimated by sentiment analysis can be viewed as a noisy signal of true underlying emotion. Due to the addition of μ_{it} into the error term, the estimates will be less efficient but remain unbiased. The task of the researcher is then simply to collect as much data as possible.

If the measurement error is suspected to be correlated with the right-hand side variable, however, biased estimates are possible. Since the measurement error is determined by the words coded for emotion, researchers are well-advised to avoid words that are likely to be dependent on temperature, such as "hot" or "cold", or even actions that are more likely to be observed differentially across temperature. The psycholinguistic literature follows this in their focus on emotional adjectives. A second implication has to do with the inclusion of I in the model: different ideas are highly likely to induce different word choices. Researchers should, to the extent possible, control for subject choice in their analyses as a method of avoiding bias. A final implication is that sorting into temperature regimes should be a concern of modelers: if individuals with different average moods choose to tweet in different temperatures, then the result could reflect a compositional change in who is expressing emotion rather than actual changes. I attempt to control for all of these possibilities in the empirical section of this paper.

A toy example: suppose that "turtles" has been assigned a low emotional score: on average, people use the word "turtles" while in depressive states. Suppose also that turtles are more frequently observed in higher temperatures. Even if there is no relationship between emotion and temperature, the frequency of the word turtles will be higher in higher temperatures for reasons having nothing to do with additional low affective states, i.e. $E[\mu_{it}T_{it}|\phi_i, \phi_t] > 0$.

2 Data

2.1 Word lists

2.1.1 Expert measure word and score samples (AFINN-111 word list)

Table 1: Expert word-score mapping examples

Positive Affect		Neutral Affect		Negative Affect	
superb	5	combat	-1	betraying	-3
thrilled	5	apologizes	-1	agonises	-3
hurrah	5	exposing	-1	destroying	-3
outstanding	5	oxymoron	-1	swindle	-3
breathtaking	5	provoked	-1	abhors	-3
roflcopter	4	limited	-1	humiliation	-3
wowow	4	escape	-1	chastises	-3
rejoicing	4	unconfirmed	-1	victimizing	-3
lifesaver	4	passively	-1	bribe	-3
winner	4	blocks	-1	lunatic	-3
miracle	4	poverty	-1	scandal	-3
triumph	4	attacked	-1	outrage	-3
wooo	4	reject	-1	haters	-3
overjoyed	4	combats	-1	assassination	-3
rejoices	4	disabling	-1	ranter	-3
rotflol	4	underestimating	-1	deception	-3
wonderful	4	banish	-1	lunatics	-3
amazing	4	crush	-1	apathetic	-3
win	4	darkness	-1	racism	-3
wowww	4	litigation	-1	swindling	-3
fun	4	drop	-1	angry	-3
awesome	4	shoot	-1	heartbreaking	-3
fabulous	4	gun	1	betrayed	-3
roflmao	4	feeling	1	terror	-3
euphoric	4	intrigues	1	abuse	-3
heavenly	4	alive	1	greenwash	-3
fantastic	4	protected	1	falsified	-3
ecstatic	4	unified	1	douche	-3
funnier	4	relieves	1	agonized	-3
winning	4	fit	1	criminals	-3
masterpiece	4	restore	1	defects	-3
masterpieces	4	relieve	1	idiotic	-3
stunning	4	greeting	1	woeful	-3
godsend	4	yeah	1	acrimonious	-3
lmfao	4	cool	1	nuts	-3
lmao	4	vested	1	swindles	-3
rotflmfao	4	clearly	1	lost	-3

Notes: Raw scores shown. Standardized scores used in analysis. Full list includes 2,478 total word-score mappings.

2.1.2 Crowd-sourced measure word and score samples (hedonometer.org word list)

Table 2: Crowd-sourced word-score mapping examples

Positive Affect		Neutral Affect		Negative Affect	
laughter	8.5	fui	5.08	suicide	1.3
happiness	8.44	gilbert	5.08	terrorist	1.3
love	8.42	hart	5.08	rape	1.44
happy	8.3	hij	5.08	murder	1.48
laughed	8.26	hun	5.08	terrorism	1.48
laugh	8.22	indonesia	5.08	cancer	1.54
laughing	8.2	jo	5.08	death	1.54
excellent	8.18	john	5.08	died	1.56
laughs	8.18	juan	5.08	kill	1.56
joy	8.16	knee	5.08	killed	1.56
successful	8.16	laws	5.08	torture	1.58
win	8.12	listed	5.08	arrested	1.64
rainbow	8.1	manhasset	5.08	deaths	1.64
smile	8.1	marion	5.08	raped	1.64
won	8.1	martinez	5.08	killing	1.7
pleasure	8.08	medicaid	5.08	die	1.74
smiled	8.08	medicine	5.08	jail	1.76
rainbows	8.06	meyer	5.08	terror	1.76
winning	8.04	might	5.08	kills	1.78
celebration	8.02	morgen	5.08	fatal	1.8
enjoyed	8.02	morris	5.08	killings	1.8
healthy	8.02	nas	5.08	murdered	1.8
music	8.02	necessarily	5.08	war	1.8
celebrating	8	norman	5.08	tortured	1.82
congratulations	8	noted	5.08	abused	1.83
weekend	8	occasionally	5.08	cruel	1.84
celebrate	7.98	ohhh	5.08	cry	1.84
comedy	7.98	ooo	5.08	failed	1.84
jokes	7.98	para	5.08	sickness	1.84
rich	7.98	pls	5.08	violence	1.86
victory	7.98	quiere	5.08	abuse	1.88
christmas	7.96	requirement	5.08	diseases	1.88
free	7.96	schemes	5.08	sadness	1.88
friendship	7.96	scott	5.08	depressing	1.9
fun	7.96	seconds	5.08	earthquake	1.9
holidays	7.96	sen	5.08	evil	1.9
loved	7.96	sets	5.08	wars	1.9

Notes: Raw scores shown. Standardized scores used in analysis. Full list includes 10,223 total word-score mappings.

2.1.3 Emoticons used for emoticon measure

Table 3: Emoticons

Positive	Affect	Negative	Affect
:-)	1	>:[	0
:)	1	:-(	0
:D	1	:(	0
:o)	1	:-c	0
:]	1	:c	0
:3	1	:-<	0
:c)	1	:?C	0
:>	1	:<	0
=]	1	:-[	0
8)	1	:[	0
=)	1	:{	0
:}	1	:'-(	0
:^)	1	:'(	0
:?)	1		
:-D	1		
8-D	1		
8D	1		
x-D	1		
xD	1		
X-D	1		
XD	1		
=-D	1		
=D	1		
=-3	1		
=3	1		
B^D	1		
:-))	1		
:))	1		

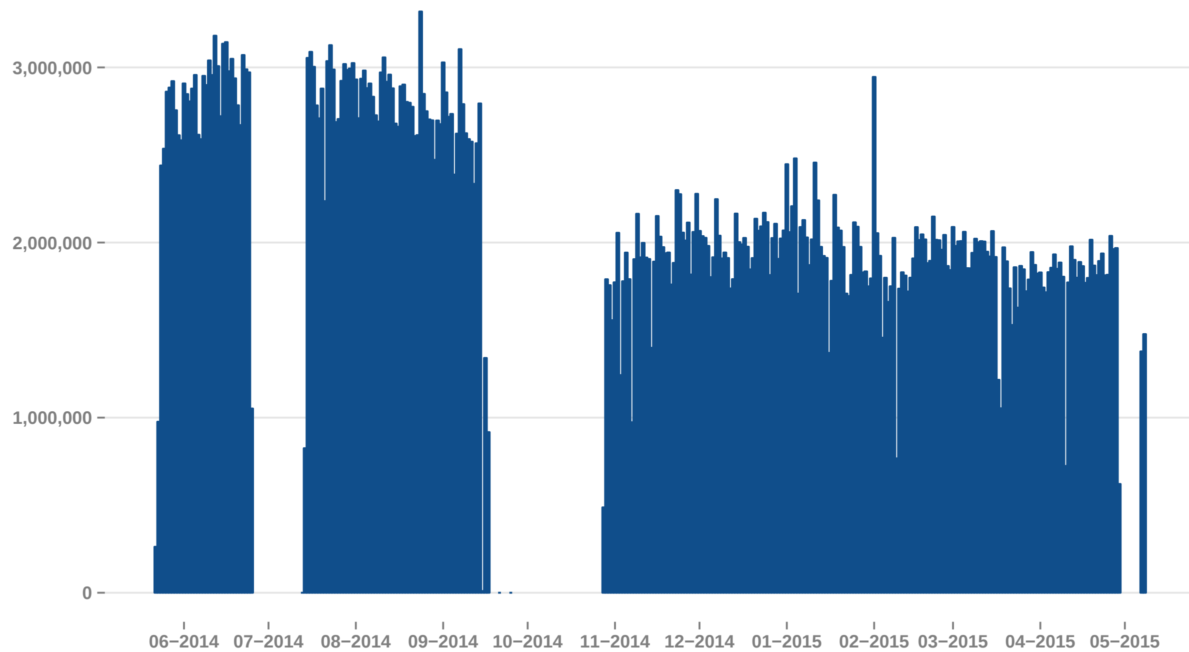
2.2 Sample of Tweets with sentiment scores

Table 4: Sample of Tweets with sentiment scores

Tweet (truncated at 50 characters)	Expert	Crowd-sourced	Profanity	Emoticon
@ThnksFrThFranks i hate you	-3.0	4.8	0	0
Just when I thought my world was crashing this lit	3.0	5.6	0	1
Killing me softly	-3.0	4.9	0	0
Savoy - Awesome EDM artists. #Osheaga2014 http://t	4.0	7.2	0	1
one thing I can't stand is a relentless liar	-2.0	4.9	0	1
Last of its kind? #walk #dontwalk #chicago http://	2.0	5.2	0	1
#Halloweengoodness #Autumnscents are 2nd best behi	3.0	5.8	0	1
@hayleeekayy too bad, I just did	-3.0	5.0	0	0
@__alexandrajoyy wtf	-4.0	4.2	0	0
All you need is love... Way to go buffet palace -	2.7	6.2	0	1
She might say she love me she don't love me like s	2.8	6.4	0	1
last shift serving at the mustang.. how weird	-2.0	4.8	0	1
Badger alumni telling me where they fell in love #	3.0	5.7	0	1
@__32keef lazy.	-1.0	3.3	0	1
Knowing ima have a chill night with bp and good pe	3.0	5.7	0	0
I would rather eat Taco Bell than eat chipotle wit	-2.5	5.4	0	0
Yeah up http://t.co/5UhuamxDQm	1.0	6.0	0	1
@Pnastyyy29 exactly lmao	4.0	5.9	0	1
#Weston, #WV for the day at the #TransAlleghenyLun	-3.0	5.2	0	1
Feeling straight up lazy today....	0.3	5.5	0	1
I'd really like it if @instagram had lists. Mostly	2.0	5.4	0	1
If she sassy we might get nasty !	-3.0	5.1	0	1
she doesn't smile when she rolls Fucking @Naydizz	-1.0	5.6	1	1
arguing with stupid people is so hard	-1.5	4.6	0	1
@__RomanL thank you	2.0	6.8	0	1
Home alone all day	-2.0	5.7	0	1
Jesus lara eres gran jugador y ojala pronto te pod	1.0	5.2	0	0
Sad I'm not going cause I wanna dance and hear dop	-1.3	5.2	1	0
Last night was perfect	3.0	5.5	0	1
Put it off long enough. I'm installing Steam on th	-4.0	5.2	1	1
@aimee__nicholee love you more	3.0	7.0	0	1
@aliyyahLENAY I love you more	3.0	6.7	0	1
@thegranddavid: Weird weather tbh	-2.0	5.0	0	0
@FontaineFairy @djfountain as you should love your	3.0	6.2	0	1
I'm so thankful for you! @ Nashville, TN http://t	2.0	5.9	0	1
Yup thats you . "@Jforde__1028: Feeling a little di	-0.5	5.3	0	1
@CrawfordCollins I promise I'm not a stalker..	2.0	5.9	0	1
Exhausted.	-2.0	3.1	0	0
@moniquearretche STFU IT FUCKING LOOKS LIKE HER	-1.0	5.7	1	1
@KendalGabrieI lol you fell asleep really early la	3.0	5.5	0	0
@KenzieStrasburg I did too and I was in heaven. He	2.0	5.8	0	1
@CaitlynMarie__16 do you still want to walk your do	1.0	5.5	0	0
5 out of 5 stars. Any other ranking is treason. -	-3.0	5.5	0	1
@jenxmartin I'll leave now too text me when you're	-1.0	5.4	0	1
Them Hittas Only Couple Of Em Left But Don't Thin	-3.0	4.8	0	0
@victoriap92 it's GONNA be so dead what time do u	-3.0	5.0	0	1
I really hate golf, how can you hit 3 shots out of	-3.0	5.0	0	1
Blank will always be one of my favorites #IMissYou	2.0	5.6	0	1
@ethanrih: Not Beyonce. Probably Katy now. "@QUEEN	2.0	5.5	0	1
Gandhi LOL @ScottGandhi	3.0	6.8	0	1
I'm having a lot of fun looking at all these fake	0.5	5.5	0	1
@HiltonSuggests thanks	2.0	7.4	0	1
I'm tired	-2.0	4.5	0	0
Everyone at buckle is so sweet to me every time I	2.0	5.8	0	1
Damn cnn is very biased hard to sit through	-2.3	4.9	1	1
Fuck ICP, buy my new CD.	-4.0	5.9	1	1
But you prolly dint remember last night no way so.	-1.0	5.0	0	0
Would anyone like to accompany me to the outlets?	2.0	5.7	0	1
Wanna do something adventurous	2.0	5.4	0	1
the heb brand cereal is better than name brand tbh	2.0	5.5	0	1
Lmao ask me how many fucks I give #YouAreIrrelevan	4.0	5.8	1	1
Pretty much only see my cousin at highland, the ga	1.0	5.6	0	1
hey twitter what bike out of these 3 do u like the	2.5	5.6	0	1
I love to travel	3.0	6.6	0	1

2.3 Tweets by date

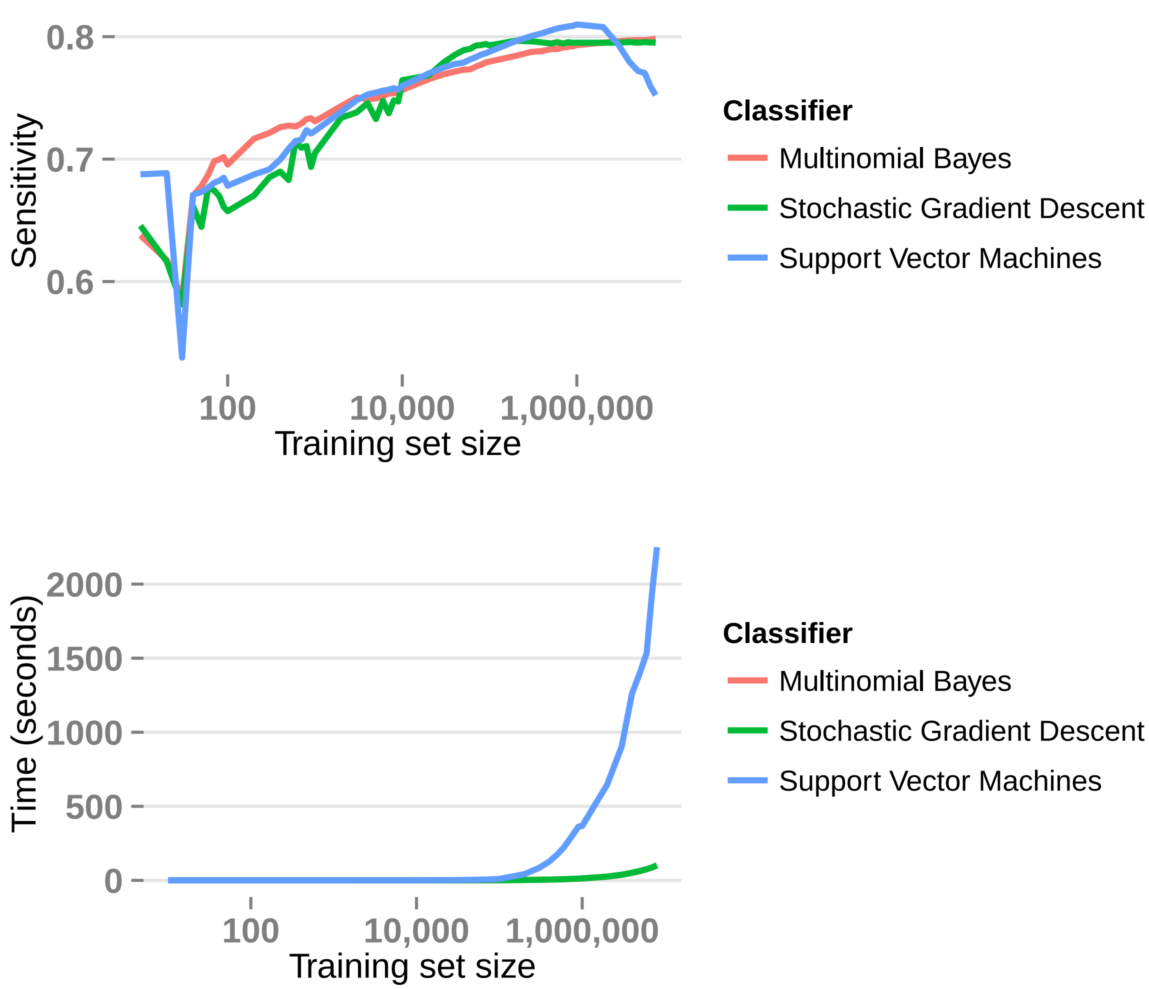
Figure 1: Tweets by date



Note: Height of bars is the number of tweets in a day. Gaps reflect two extended disconnections from Twitter API.

2.4 Comparison of Machine Learning techniques

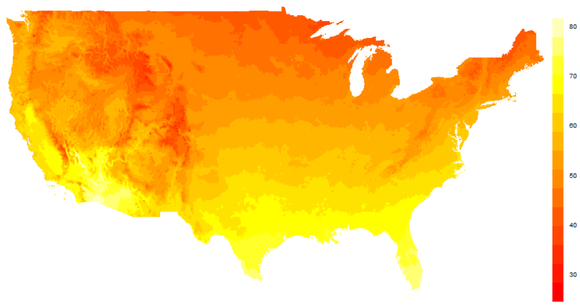
Figure 2: Comparison of machine learning techniques



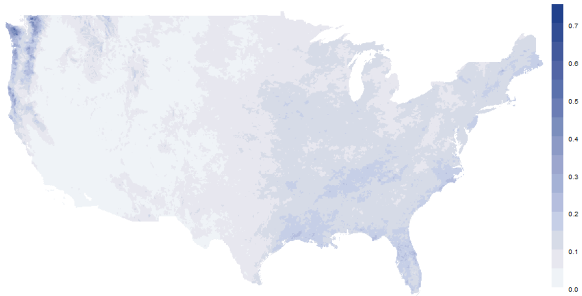
2.5 Weather data maps

2.5.1 PRISM

Figure 3: PRISM sample frame means



(a) Mean average daily temperature



(b) Mean rainfall

Note: Top map is mean average temperature in °F experienced by grid cell during sample frame. Bottom map is mean daily rainfall.
Source: PRISM data.

2.5.2 QCLCD

3 Empirics

3.1 Five degree bins

3.2 Cubic splines

3.3 Minimum temperature

3.4 Maximum temperature

3.5 Regression tables for Crowd-sourced and Profanity measures

Table 5: Effect of temperature on hedonic state (Crowd-sourced measure)

	(1)	(2)	(3)	(4)	(5)	(6)
Daily temperature T						
$T < 20$	–0.007	0.002	0.007	0.006	0.006	0.002
	(0.005)	(0.004)	(0.008)	(0.008)	(0.007)	(0.006)
$T \in [20, 30)$	–0.009**	0.001	0.005	0.004	0.005	0.001
	(0.004)	(0.004)	(0.007)	(0.007)	(0.006)	(0.006)
$T \in [30, 40)$	–0.009*	0.001	0.004	0.003	0.004	0.002
	(0.004)	(0.004)	(0.007)	(0.007)	(0.005)	(0.005)
$T \in [40, 50)$	–0.005	0.004	0.005	0.006	0.006	0.004
	(0.004)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
$T \in [50, 60)$	–0.000	0.001	0.002	0.002	0.002	0.002
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
$T \in [70, 80)$	–0.014***	–0.006***	–0.007***	–0.005***	–0.005***	–0.004***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$T \in [80, 90)$	–0.033***	–0.011***	–0.012***	–0.009***	–0.010***	–0.008***
	(0.003)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
$T \geq 90$	0.000	–0.013***	–0.013***	–0.011***	–0.015***	–0.014***
	(0.006)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Grid cell-days (m.)	23.1	23.1	23.1	23.01	22.25	22.25
Twitter updates (m.)	871.2	871.2	871.2	871.2	846.8	846.8
County FE	No	Yes	Yes	No	No	No
Grid cell FE	No	No	No	Yes	Yes	Yes
Month FE	No	Yes	No	No	No	No
State×Month FE	No	No	Yes	Yes	Yes	Yes
Day of week FE	No	No	No	No	No	Yes
Weather controls	No	No	No	No	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Dependent variable is the mean Crowd-sourced measure of hedonic state for a grid cell-day, standardized to have zero mean and unit standard deviation. Independent variables are dummies for temperature (in degrees F) bins. Each column is a separate regression, coefficients represent the mean difference in standardized hedonic state between a day with the associated temperature bounds relative to a day with temperature $T \in [60, 70)$, the omitted category. Coefficients are estimated conditional on the controls listed in the bottom half of the table. Clustered two standard errors on county by month of sample and day of sample in parentheses.

Table 6: Effect of temperature on hedonic state (Profanity measure)

	(1)	(2)	(3)	(4)	(5)	(6)
Daily temperature T						
$T < 20$	–0.005	–0.004	0.000	–0.001	–0.002	–0.003
	(0.005)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
$T \in [20, 30)$	–0.006	–0.002	0.003	0.001	0.001	–0.001
	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$T \in [30, 40)$	–0.009**	–0.002	0.002	0.000	0.001	–0.001
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
$T \in [40, 50)$	–0.007*	–0.002	0.001	0.001	0.001	0.000
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$T \in [50, 60)$	0.003	–0.001	0.000	0.000	0.000	0.000
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$T \in [70, 80)$	–0.014***	–0.006***	–0.006***	–0.004***	–0.004***	–0.003***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
$T \in [80, 90)$	–0.031***	–0.010***	–0.010***	–0.007***	–0.007***	–0.006***
	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$T \geq 90$	–0.004	–0.013***	–0.011***	–0.009***	–0.010***	–0.009***
	(0.008)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Grid cell-days (m.)	23.46	23.46	23.46	23.36	22.59	22.59
Twitter updates (m.)	945.2	945.2	945.2	945.2	918.9	918.9
County FE	No	Yes	Yes	No	No	No
Grid cell FE	No	No	No	Yes	Yes	Yes
Month FE	No	Yes	No	No	No	No
State×Month FE	No	No	Yes	Yes	Yes	Yes
Day of week FE	No	No	No	No	No	Yes
Weather controls	No	No	No	No	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Dependent variable is the mean Profanity measure of hedonic state for a grid cell-day, standardized to have zero mean and unit standard deviation. Independent variables are dummies for temperature (in degrees F) bins. Each column is a separate regression, coefficients represent the mean difference in standardized hedonic state between a day with the associated temperature bounds relative to a day with temperature $T \in [60, 70)$, the omitted category. Coefficients are estimated conditional on the controls listed in the bottom half of the table. Clustered two standard errors on county by month of sample and day of sample in parentheses.

3.6 Estimates with user fixed effects for Crowd-sourced and Profanity measures

3.7 User $\times$ grid cell fixed effects

3.8 Winter heterogeneity for Crowd-sourced and Profanity measures

3.9 Decomposition of variance

3.10 Projection methodology detail

To project the effects of climate change on hedonic state, I first take the average predicted by-month temperature change across an ensemble of models from the Coupled Model Intercomparison Project 5 (WRCP 2011), downscaled using the Multivariate Adaptive Constructed Analogs (Abatzoglou and Brown 2012) method. To avoid aggregation bias, I follow the recommendations of Auffhammer et al. (2013): I add the differences between the model averages of months in the 2000-2019 projection and the 2086-2099 projections to historical temperature data. Next, I estimate the annual average number of days with average temperature in the bins defined in model (**??**) under the Representative Concentration Pathway 4.5 (RCP4.5) and 8.5 (RCP8.5) scenarios.[4] I then multiply the change in the annual average days in a temperature bin with the change in hedonic state relative to the 60-70°F bin. Taking the sum over these products gives the average annual change in hedonic state in standard deviations that would be induced by the projected end-of-century changes in climate. Because I use downscaled climate data, I am able estimate the annual change separately for every grid cell in my data set. The results are qualitatively similar across all measures, while the changes under RCP4.5 are less pronounced than those under RCP8.5, as expected. Darker areas indicate larger (in absolute value) levels of damages. In general, differences reflect changes in the temperature distribution: since the climate models predict on average high temperatures in the southern areas of the United States, these areas are also predicted to see larger changes in hedonic state. This thought experiment implicitly holds technology and economic growth constant and does not include the effects of long-run adaptation.

The density of my data allow me to combine the analysis conducted here with that of the prior section to more precisely account for geographic differences in temperature response functions. To do so, I estimate five separate response functions for each quintile of historical daily temperature in my sample. These response functions implicitly allow for long-run

[4]RCP scenarios capture different levels of radiative forcing, or emissions, where RCP4.5 represents a low-emissions pathway and RCP8.5 represents a high-emissions pathway.

adaptation to temperature changes. As in the above section, I note that the magnitude of the response to warmer temperatures decreases for hotter regions. Next, I conduct the following projection exercises: first, I conduct the same exercise as above using the disaggregated quintile response functions, holding constant the response function for each geographical location. Second, I allow grid cells to adapt to changing average temperature, adding the difference between the temperature in either the RCP4.5 or the RCP8.5 scenarios to the historical average temperature for that grid cell. Based on the sum of the historical average temperature and the increase in temperature predicted by the forcings scenario, I project the estimated damages for each grid cell *using the response function for the corresponding quintile.*

3.11 Response functions by historical temperature quintile

3.12 Precipitation

I allow precipitation to enter the regression flexibly in 1 millimeter bins, with the omitted category being days on which no precipitation fell.

$$\overline{\hat{E}}_{gd} = \alpha + \sum_{b \neq 0}^{B} \beta_b P_{gd}^{b} + \phi_g + \phi_{sm} + \varepsilon_{gd}$$

I estimate a slight negative effect of precipitation on hedonic state, finding magnitudes of around 0.005σ on average across sentiment measures. I am able to reject the null hypothesis of a positive effect of rainfall on human mood. These findings do not confirm the validity of rainfall as an instrument, but they suggest that changes in hedonic state are not a likely mechanism.

3.13 Tweet frequency by temperature for high and low sentiment users

I also propose a more direct method to test whether or not high and low sentiment users tweet differently is to examine the how tweet frequency changes with temperature for groups with high or low average sentiment. I separate the users into two groups: users with mean sentiment (as measured by the Expert score) in the 40th percentile or lower and users with mean sentiment in the 60th percentile or higher; users in the middle are dropped. I then estimate tweet frequency for each group of users. Figure **??** shows the results: the frequency of status updates does not appear to vary with high or low mean sentiment.

3.14 Lagged effects of temperature

A common feature of relationships between weather and economic outcomes is that there is a lag structure to the effects. In the case of affect, if affect is as sensitive to temperature yesterday or the day before then it could suggest that some other factor is driving the effects estimated by specification **??**. To test this, I generate estimates using zero, one and two day lags for rows 1-3 respectively in Figure **??**. While the contemporaneous effect is stable whether or not lags are included, the lagged estimates do not tend to be significantly different from zero, although there is some limited evidence of a positive relationship between affect and lagged weather.

References

Abatzoglou, John T., and Timothy J. Brown. 2012. "A comparison of statistical downscaling methods suited for wildfire applications". *International Journal of Climatology* 32 (5): 772–780.

Auffhammer, Maximilian, Solomon M. Hsiang, Wolfram Schlenker, and Adam Sobel. 2013. "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change". *Review of Environmental Economics and Policy* 7 (2): 181–198.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter mood predicts the stock market". *Journal of Computational Science* 2 (1): 1–8.

Clark, Herbert H. 1975. *Speech errors as linguistic evidence.*

Dodds, Peter Sheridan, and Christopher M. Danforth. 2010. "Measuring the happiness of large-scale written expression: Songs, blogs, and presidents". *Journal of Happiness Studies* 11 (4): 441–456.

Edgeworth, Francis Ysidro. 1881. "Mathematical psychics". *Mind* 6:581–583.

Garrett, M.F. 1976. "Syntactic process in sentence production". In *New Approaches to Language Mechanisms*, 30:231–256.

Gerber, Matthew S. 2014. "Predicting crime using Twitter and kernel density estimation". *Decision Support Systems* 61 (1): 115–125.

Mitchell, Lewis, Morgan R. Frank, Kameron Decker Harris, Peter Sheridan Dodds, and Christopher M. Danforth. 2013. "The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place". *PLoS ONE* 8 (5).

Pak, Alexander, and Patrick Paroubek. 2010. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." In *Lrec*, 10:1320–1326.

Pang, Bo, and Lillian Lee. 2008. *Opinion Mining and Sentiment Analysis.*

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* 58:752. 2.

WRCP. 2011. “WCRP Coupled Model Intercomparison Project”. *CLIVAR Exchanges* 16 (56).