Recreation of Temperature and Decisions: Evidence from 207,000 Court Cases

Viktor Reif

January 31, 2022

Abstract

I replicate the main results of the paper "Temperature and Decisions: Evidence from 207,000 Court Cases" by Heyes and Saberian (2019) and evaluate the validity of its empirical findings. The paper's underlying specification to model the effect of outdoor temperature on asylum court case outcomes is a panel model including control variables for weather, pollution and case characteristics as well as several fixed effects for entities, time and space dynamics. I find the same effects as Heyes and Saberian. Yet through a more thorough analysis (sampling, specification and data issues) get results that question the validity of the original findings. Using yearly subsamples of the entire dataset (2000-2004) yields temperature effects that are not significantly different from zero for all but one year.
Keywords: decision-making; temperature; panel data regression; latent factors

1 Introduction

This paper examines the robustness of the results in the article "Temperature and Decisions: Evidence from 207,000" by Heyes and Saberian in 2019. Using the same dataset, this paper recreates the main findings. The aim of this paper is to either empirically confirm the results of Heyes and Saberian or to disprove them and illuminate the reasons for that. I analyse the dataset in Python, recreate the most relevant tables from the original paper in Stata and implement an example for a more sophisticated specification in R.

Heyes and Saberian (2019) use a dataset of 207,000 court cases in the U.S. for a holistic regression analysis to evaluate the influence of outdoor temperature on professionally made decisions. The authors use a large set of explanatory variables including various fixed effects - over time, across judges and locations, etc - to control for heterogeneity in the regression of court case outcome on temperature. In this analysis they find a significant relation between temperature and likeliness that a case has a positive outcome, meaning that asylum is granted.

I am able to replicate the paper's main finding. In my analysis, estimated coefficients are equal in value, direction and significance. Moreover, I find that the underlying analysis has some (minor) data issues, which if taken care of, lead to results that do not contradict the main finding. A more critical insight on my side was that taking subsamples of the dataset yields insignificant temperature effects. This finding strongly questions the validity of the original results.

2 Literature review

Heyes and Saberian already give an exhaustive overview in their paper from 2019. Their work is in line with numerous publications showing that temperature - both indoors and outside - does have a significant effect on human decisions and rationality. More recently, in this branch of literature that finds a relationship between temperature and decision making, Gavresi et al. (2021) show that higher outdoor temperature increases risk appetite in (optimist) financial decisions. Chen et al. (2020) find that people perform worse in neurobehavioral cognitive tests when exposed to higher temperature indoors and Hadi and Block (2019) show that extreme heat makes consumers less rational (ie affectual). Even more temperature effects are shown by Stevens et al. (2021) on aggression on social media and by Ryan (2020) on law officials behaviour.

There is also a group of researchers who disprove the link between temperature and decisions, which Heyes and Saberian omit in their paper. Recent contributions in this branch are Stroom et al. (2021), who find no relation between indoor temperature and cognitive rationality, and Liu et al. (2020), who observe no effect of heat on fraudulent behaviour. Contributions before 2019 in that direction are Zhang and de Dear (2017) and Zhang et al. (2017). The former finds no indoor temperature (excluding extreme conditions) effects on university student performance and in the latter office workers' cognitive load was not significantly affected by room temperature.

Concerning temperature effects on juridical outcomes specifically, Heyes and Saberian are the first to conduct a full empirical analysis. This motivated Evans and Siminski (2021) to do their own empirical analysis about criminal court cases in Australia, which resulted in no significant effect between weather variables and decision making. Also, as direct response to the underyling paper Holger Spamann (2020) recalculates its results within a larger timeframe (1990 - 2019) and finds no significant effects.

3 Data

The main dataset is constructed out of several sources. asylumlaw.org contains the law data in the form of the variables case outcome, case type and nationality of applicant structured along the dimensions judge, city and date (No combinations of those builds unique keys). For the environment data, the National Oceanic and Atmospheric Administration yields air temperature, dew point, air pressure, precipitation and wind speed sorted hourly by datetime and location. The variable cloud cover is available at the Northeast Regional Climate Center. The pollution variables quantity of micro particles, carbon monoxide and ozone are delivered by United States Environmental Protection Agency. Some of the environment data is collected hourly and some daily. As the law variables are in a daily format, hourly data is averaged daily from 6AM to 4PM. Each environment observation is at maximum 32 kilometers away from the respective court location. All variables are joined by date (daily) and city in the dataset matched.dta, such that every row represents one case outcome marked by a respective date and location enhanced by case-related as well as environmental characteristics.
Once matched.dta is created by joining all data sources, it contains 269269 observations for 572 characteristics. The (Stata) code then converts certain temperature variables into promils and creates auxiliary variables for all relevant dimensions (city, judge, year, month, day), averages of some characteristics across various dimensions and interactions between variables. The final dataset contains 206924 usable observations for 588 variables (the model uses 6 specific non-numeric characteristics as fixed effects, for which Stata will temporarily create a total of 1006 dummy columns within the regression, thus that the total number of variables will be 1596).

Table 1: Summary Statistics

	Mean	Std. Dev.
res	0.165	0.371
tempmean	61.577	15.124
heat	57.642	16.393
airpressure0	29.657	0.780
avgdewpt	50.116	16.777
precip0	0.004	0.039
windspeed0	6.655	4.434
skycover	0.556	0.276
ozone	0.022	0.012
co	0.916	0.496
pm25	14.751	11.374

Table 1 shows summary statistics for the most relevant variables. About 16 percent of all cases end in granting the applicant asylum (represented by the variable ”res”). As noted by Heyes and Saberian, the grant rate differs greatly across judges and location. For instance over the study period in the Los Angeles courthouse there are five judges that granted asylum to fewer than 4 percent while three others granted in over 67 percent. The mean over the entire dataset for daily average temperature is 61.4°F, which is around 14°C.
Figure B.1 (appendix) shows the distribution of NA values across variables in the dataset. Whereas all variables show rather complete observations, the variable ”co” (carbon monoxide) contains approx. 50000 missing values. That issue is addressed in Stata within the regression by dropping every row that contains NA for at least one included variable. The variable carbon monoxide is here especially noteworthy, as its missing values cover all data from the year 2001. Subsequently, if ”co” is used as an explanatory variable in a regression, Stata will drop all observations from 2001 for the estimation.

To illustrate this loss of data, Figure B.2 shows on the left-hand side the yearly observation count of the dataset without any NA values dropped and on the right-hand side the same count with carbon monoxide NA values dropped. Note that excluding "co" as a regressor sets the effective number of observations to 250651.

4 Empirical strategy

The main hypothesis to be tested is whether outdoor temperature has an impact on professional high-stakes decisions. In a more empiric logic, this hypothesis is tested using a linear probability model for binary response estimated by Pooled Ordinary Least Squares (for a detailed description see Wooldridge (2010)). The probability model allows each regressor to influence the likelihood that the dependent variable takes the value of 1. A value of 1 means that asylum is granted. The following model tests the main hypothesis,

$$g_{it} = \beta_0 + \beta_1 temp_{it} + W_{it}\beta_2 + P_{it}\beta_3 + X_{it}\beta_4 + \gamma_i + \psi_{ct} + \theta_t + \epsilon_{it} \tag{1}$$

where the dimensions i, t and c represent application, date and city respectively. Thus, the regressand g_{it} is the outcome of an application i on the date t having the value 1 if asylum was granted and 0 otherwise. β_j are the j slope parameters the regressors (or element of respective regressor matrix) and β_0 is used as the intercept. *temp* is the main regressor of interest, whereas W_{it}, P_{it} and X_{it} are a set of control variables. W_{it} includes averaged weather characteristics (skycover, air pressure, wind, precipitation, dewpoint), P_{it} represents air pollution (ozone, co, pm) and X_{it} includes all case specific dummies (weekday, nationality of applicant, case type, year, city-month interaction). γ_i, ψ_{ct} and θ_t are included to control for judge-specific fixed effects (ie which judge is ruling the case), time fixed effects (weekday and years) and city-by-month effects. ϵ_{it} contains unobserved heterogeneity along the dimensions of case and date. This serves to control for time and spatial autocorrelation. The fixed effects regression model is especially suitable for this analysis for two reasons. Firstly, it yields a handy interpretation for the coefficient of interest, which is in turn also comparable to several other studies, that used the same approach to model decision making or temperature effects. As this is a probability model, the effect on the dependent variable will always be a change in likelihood (%). Moreover, it is easy to determine and interpret significance of all controls and fixed effects. Secondly, this model can include many characteristics fixed in their respective dimensions. Thusly, as done by Heyes and Saberian, the fixed effects model allows for a holistic approach when testing and including numerous fixed effects.

Apart from the normal specification, I also propose a slightly altered version of (1), in which "co" is excluded from P_{it} to account for the missing value issue explained in the Data chapter. As carbon monoxide is one of several pollution controls, its exclusion should not cause a strong omitted variable bias. Moreover, I estimate (1) with yearly subsamples of the dataset to evaluate the model's validity in a smaller sampling scope. This yearly analysis is motivated by the findings from Holger Spamann (2020). The intuition is that if a larger scope yields results, that question the original finding, then using a smaller scope might also find relevant differences.

As an alternative specification, which extends (1), I propose the inclusion of latent factors. This is helpful if the model is underspecified due to relevant variables not being available. Here, one feasible omitted variable could be for instance judges' general mood and feeling of well-being, which arguably changes across judges, over time or spatially and influences asylum case outcomes. This regressor - or any other - can still be controlled for by deducting latent factors from the error term. In that way less biased estimates can be achieved, even though is it unknown which characteristics each of these factors represent. The following is the latent factor model

$$g_{it} = \beta_0 + \beta_1 temp_{it} + W_{it}\beta_2 + P_{it}\beta_3 + X_{it}\beta_4 + \gamma_i + \psi_{ct} + \theta_t + \upsilon_{mn} + \epsilon_{it} \tag{2}$$

, where

$$v_{mn} = \sum_{l=1}^{d} \lambda_{ml} f_{ln} \tag{3}$$

are interactive fixed effects structured along the dimensions mn, which are arbitrary but usually chosen to follow the dimensions of the regressand (it). As the dataset lacks unique key combinations, I specify arbitrary ranges for mn (see R code for details). λ_{ml} are unobserved individual loading parameters. f_{ln} represent unobserved common factors of the model and d is the unknown factor dimension. It is important to notice that λ_{ml} and f_{ln} are unobserved, because they are principal components (multidimensionally orthogonal vectors as a result from an eigenvalue decomposition) drawn from the error term. As in (1), the model treats the parameters of the regressors and the unobserved effects as fixed coefficients and estimates them. As the unobserved effects are not part of the error term, they are allowed to correlate with the regressors in any possible way and thusly influence their estimated parameters (See Bai (2009) for details).

5 Results

Table 2 contains the results of the regression using the default specification. All four regressions use pooled OLS to estimate the effects of average temperature and its one-day lag as well as lead in different combinations. Also, all specifications control for a set of averaged weather characteristics, air pollution and case specific characteristics. In column (1) of Table 2, the estimated slope parameter is -1.075. This value means that a 10°F (5.4°C) increase in daily average temperature during a judge decision reduces the probability of a positive outcome by 1.075% (ceterus paribus). Considering that the overall average grant rate is 16.3%, a 10°F warmer temperature implies a 6.59% decrease in expected grant rate. This effect is statistically significant at 1%. In (2), (3) and (4) the same effect has different values but remains equal in direction and significance. Analogously to the just interpreted parameter, the lag or lead estimates quantify the effect of the average temperature the day before or after the decision. In no specification the regression analysis finds lead or lag effects significantly different from zero. This means that in this dataset the outdoor temperatures the day after and before a court decision is made have no effect on its outcome.

Table 2: Fixed effect estimates: 6 AM - 4 PM average

	(1) base	(2) 1-Daylag	(3) 1-Day lead	(4) all
temp6t410	-1.075***	-1.454***	-1.208***	-1.617***
	[0.274]	[0.406]	[0.382]	[0.486]
press6t4	-0.00494	-0.00500	-0.00515	-0.00523
	[0.00518]	[0.00518]	[0.00516]	[0.00516]
dew6t4	0.000723***	0.000765***	0.000735***	0.000780***
	[0.000213]	[0.000217]	[0.000217]	[0.000222]
prcp6t4	0.0616	0.0590	0.0625	0.0600
	[0.0822]	[0.0821]	[0.0820]	[0.0818]
wind6t4	0.000738	0.000771	0.000820	0.000866
	[0.000490]	[0.000485]	[0.000548]	[0.000543]
skycover	-0.00292	-0.00159	-0.00186	-0.000343
	[0.00501]	[0.00515]	[0.00538]	[0.00551]
ozone	0.493***	0.503***	0.485***	0.494***
	[0.160]	[0.160]	[0.157]	[0.157]
co	0.00572	0.00547	0.00552	0.00523
	[0.00389]	[0.00389]	[0.00385]	[0.00384]
pm25	-0.00000866	-0.0000104	-0.0000130	-0.0000153
	[0.0000987]	[0.0000986]	[0.000100]	[0.0000999]
ltemp6t410		0.361		0.372
		[0.278]		[0.277]
letemp6t410			0.139	0.159
			[0.260]	[0.260]
N	206924	206924	206924	206924

Standard errors in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.1 (see appendix) lists several alternative specifications and is equal to the results by Heyes and Saberian. Table A.2 shows the results of re-estimating Table 2 without ″co″ as a control variable. Temperature effects change slightly in value whereas direction and significance remain unchanged. This implies that including the observations (blocked before by the NA values in ″co″) from 2001 and omitting ″co″ as control yields similar results.

Holger Spamann (2020) have already strongly undermined the validity of the results by Heyes and Saberian using a larger scope for their sample selection and not finding significant temperature effects. Table A.3 lists the first column of the base result (ie base model without lags/leads) for yearly subsamples of the main dataset. As for the year 2001 there is no co2 available in the dataset, it is omitted for that year. Each yearly result yields negative coefficients for the relation between temperature and grant rate, yet for all years except 2003 this effect is not significant. Hence, applying a smaller scope for the sample selection also puts the original results in question. When combining the findings of Spamann with Table A.2, a potential interpretation could be that by chance Heyes and Saberian found significant effects in exactly their sample scope (2000-2004). A short side note here is that the original analysis controls for years with dummies. Those are significant, yet these fixed effects only control for yearly differences in the dependent variable between the base year (2000) and each other year. This does not control for for differences in the regressors, and therefore not control for different yearly temperature effects (as seen Table A.3) either. Table A.5 (R output) shows the estimates of (2) without any lags or leads for temperature. The dimension of the unobserved factors is two, meaning that the dimensionality criterion "PC3" found the inclusion of two latent factors is the best trade off between increase in goodness of fit and loss of degrees of freedom. Controlling for these two latent factors leads to a slightly larger temperature effect, which is still highly significant.

6 Discussion

The results of this paper are in line with the findings from Heyes and Saberian and so are its implications. The results imply that highly important decisions might not be arbitrary. These shortcomings decrease overall societal welfare and efficiency.

As noted in the literature review, other authors have shown that the data selection as done by Heyes and Saberian is flawed. As this analysis uses the same dataset, it suffers from the same obvious limitation. Another limitation might come from the model specification. In that regard there could exist omitted variables that bias the coefficient of interest either due to unavailability of data (variables) or a too parsimonious methodology. Going back to the chapter "Empirical Strategy", judges' general mood and feeling of well-being serves as a theoretical example for an omitted variable bias. Judge fixed effects as in (1) are unarguably insufficient to control for this surely also time-varying characteristic. The same issue might arise for other unobserved factors, which are all put into the error term and threaten the validity of the model. Here, the proposed latent factor model serves as an example how the data in question, or more specifically, the error term of the model can be "mined" to get better estimates. It is not guaranteed that (2) gets us closer to the true relation between temperature and decisions. There are many more approaches to improve the underlying model using the same data. Therefore the proposed latent factor model should be seen more like a short example rather than a solution to the problem described earlier. Antoher small note is that as in Heyes and Saberian, I use a linear probability model, even though this comes with the threat of fitted values falling outside of the [0; 1] interval. This threat, however, does not seem to be a relevant problem in this paper or similar literature, whilst results remain comparable with other papers.

More research on the link between temperature and decision making is needed to clarify the ongoing uncertainty about it. Especially in the context of court cases more empiric contributions would help, as those are yet scarce. Furthermore, future research can also examine the relation decisions/ human cognitive output and climate change. This is outstandingly relevant as average global temperatures as well as local weather volatility are expected to rise. The relation between decisions and climate change might even be impactful enough to be included in the integrated modelling approaches (IAMs) of global temperature forecasts. Going back to the basic relation between decisions and current weather, another topic for research would be to mitigate the methodological shortcomings mentioned above. This could be done either through mining techniques to achieve a more adequate model specification, or use other analysis/estimation methods. Examples for the latter would be regression trees or neural networks.

7 Conclusion

The main findings of this paper are twofold. One one hand the analysis can confirm the correctness of the empirical findings from Heyes and Saberian. On the other hand a more holistic analysis on my part strongly questions the validity of the underlying paper's findings. Apart from smaller issues, the effects found by Heyes and Saberian do not hold for subsamples (eg yearly intervals) of the main dataset in question. These validity issues found in sampling scopes smaller than the original scope are in line with the finding of other authors. Holger Spamann (2020) uses a larger scope (1990-2019 instead of 2000-2004) and finds no significant temperature effects. Hence, this paper questions the validity of the originally found effects from yet another empirical perspective. With those main findings this paper gravitates towards the stream of publications that negate temperature effects on decision making. Yet this paper also shows several avenues for research to further understand the relation between exogenous variables and human decisions/ cognitive outputs.

Appendix

A Additional Tables

Table A.1: Fixed effect estimates: 6 AM - 4 PM average

	(1) nothing	(2) nat	(3) dow	(4) type	(5) judge	(6) cm	(7) city/ym	(8) cym
temp	-1.470***	-0.717***	-0.727***	-0.780***	-0.806***	-1.037***	-0.893***	-0.652**
	[0.355]	[0.270]	[0.273]	[0.269]	[0.249]	[0.278]	[0.215]	[0.262]
N	206924	206924	206924	206924	206924	206924	206924	206924

	(9) jm/c/y	(10) date	(11) base
temp	-1.073***	-0.939***	-1.075***
	[0.271]	[0.285]	[0.274]
N	206924	206924	206924

Standard errors in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Fixed effect estimates: 6 AM - 4 PM average (without carbon monoxide as control variable)

	(1) base	(2) 1-Daylag	(3) 1-Day lead	(4) all
temp6t410	-0.877***	-1.237***	-1.147***	-1.532***
	[0.260]	[0.365]	[0.352]	[0.454]
press6t4	-0.00535	-0.00543	-0.00581	-0.00592
	[0.00470]	[0.00469]	[0.00465]	[0.00464]
dew6t4	0.000638***	0.000676***	0.000659***	0.000699***
	[0.000202]	[0.000206]	[0.000206]	[0.000211]
prcp6t4	0.0293	0.0270	0.0313	0.0290
	[0.0798]	[0.0796]	[0.0796]	[0.0795]
wind6t4	0.000760*	0.000805*	0.000939*	0.000994*
	[0.000458]	[0.000457]	[0.000512]	[0.000513]
skycover	-0.00646	-0.00524	-0.00426	-0.00289
	[0.00454]	[0.00455]	[0.00475]	[0.00480]
ozone	0.120	0.128	0.105	0.112
	[0.132]	[0.133]	[0.131]	[0.132]
pm25	0.0000481	0.0000451	0.0000363	0.0000325
	[0.0000965]	[0.0000963]	[0.0000980]	[0.0000979]
ltemp6t410		0.343		0.356
		[0.261]		[0.262]
letemp6t410			0.282	0.295
			[0.248]	[0.249]
N	250652	250652	250652	250652

Standard errors in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Yearly results

	(1)	(2)	(3)	(4)
	2000	2002	2003	2004
temp6t410	-0.644	-0.425	-0.928**	-0.268
	[0.648]	[0.509]	[0.431]	[0.635]
press6t4	-0.00770	0.0150	-0.00718	-0.0140
	[0.0192]	[0.0153]	[0.0151]	[0.0173]
dew6t4	-0.0000398	0.000308	0.000371	0.000829*
	[0.000453]	[0.000375]	[0.000325]	[0.000483]
prcp6t4	-0.00440	0.0848	-0.0245	0.143
	[0.157]	[0.161]	[0.128]	[0.203]
wind6t4	-0.000105	0.000106	0.00149*	0.000539
	[0.000999]	[0.000854]	[0.000827]	[0.000929]
skycover	0.00180	-0.00775	-0.00695	0.00676
	[0.0111]	[0.00905]	[0.00898]	[0.00908]
ozone	0.840***	0.566*	0.199	-0.0777
	[0.294]	[0.304]	[0.258]	[0.317]
co	0.00302	0.00172	0.00458	0.0104
	[0.00681]	[0.00791]	[0.00788]	[0.0115]
pm25	0.0000814	0.0000511	0.000270	0.000163
	[0.000131]	[0.000167]	[0.000190]	[0.000290]
N	45463	54106	65572	41783

Standard errors in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Yearly results (without carbon monoxide as control variable)

	(1)	(2)	(3)	(4)	(5)
	2000	2001	2002	2003	2004
temp6t410	-0.644	-0.109	-0.428	-0.943**	-0.240
	[0.648]	[0.675]	[0.509]	[0.428]	[0.640]
press6t4	-0.00778	0.0000866	0.0153	-0.00638	-0.0120
	[0.0193]	[0.0164]	[0.0154]	[0.0153]	[0.0174]
dew6t4	-0.0000237	0.000538	0.000317	0.000395	0.000844*
	[0.000453]	[0.000587]	[0.000371]	[0.000328]	[0.000486]
prcp6t4	-0.00413	-0.101	0.0854	-0.0228	0.151
	[0.157]	[0.209]	[0.161]	[0.128]	[0.203]
wind6t4	-0.000236	0.00130	0.0000530	0.00136*	0.000325
	[0.000896]	[0.00103]	[0.000821]	[0.000781]	[0.000924]
skycover	0.00131	-0.0186*	-0.00787	-0.00709	0.00606
	[0.0110]	[0.00980]	[0.00899]	[0.00895]	[0.00898]
ozone	0.841***	-0.459	0.553*	0.164	-0.148
	[0.293]	[0.303]	[0.293]	[0.251]	[0.309]
pm25	0.0000932	0.000141	0.0000607	0.000310*	0.000215
	[0.000129]	[0.000271]	[0.000162]	[0.000178]	[0.000283]
N	45463	43728	54106	65572	41783

Standard errors in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Latent factor estimates

```
Call:
Eup.default(formula = res ~ temp6t410 + press6t4 + dew6t4 + prcp6t4 +
    wind6t4 + skycover + ozone + co, dim.criterion = "PC3")

Residuals:
    Min      1Q  Median      3Q     Max
-0.4740 -0.1570 -0.0728  0.0437  1.0200


 Slope-Coefficients:
             Estimate   Std.Err Z value    Pr(>z)
(Intercept)  0.163000  0.001630 100.000 < 2.2e-16 ***
temp6t410   -1.360000  0.298000  -4.570  4.83e-06 ***
press6t4    -0.000608  0.002310  -0.263  0.792000
dew6t4       0.000585  0.000270   2.170  0.030200 *
prcp6t4      0.357000  0.130000   2.750  0.005940 **
wind6t4      0.002010  0.000544   3.690  0.000222 ***
skycover     0.028700  0.006410   4.490  7.24e-06 ***
ozone        1.150000  0.161000   7.110  1.13e-12 ***
co           0.019900  0.003630   5.480  4.37e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Additive Effects Type:  none

Dimension of the Unobserved Factors: 2

Residual standard error: 0.3594 on 47395 degrees of freedom,
R-squared: 0.5305
```

B Figures

Figure B.1: NA distribution

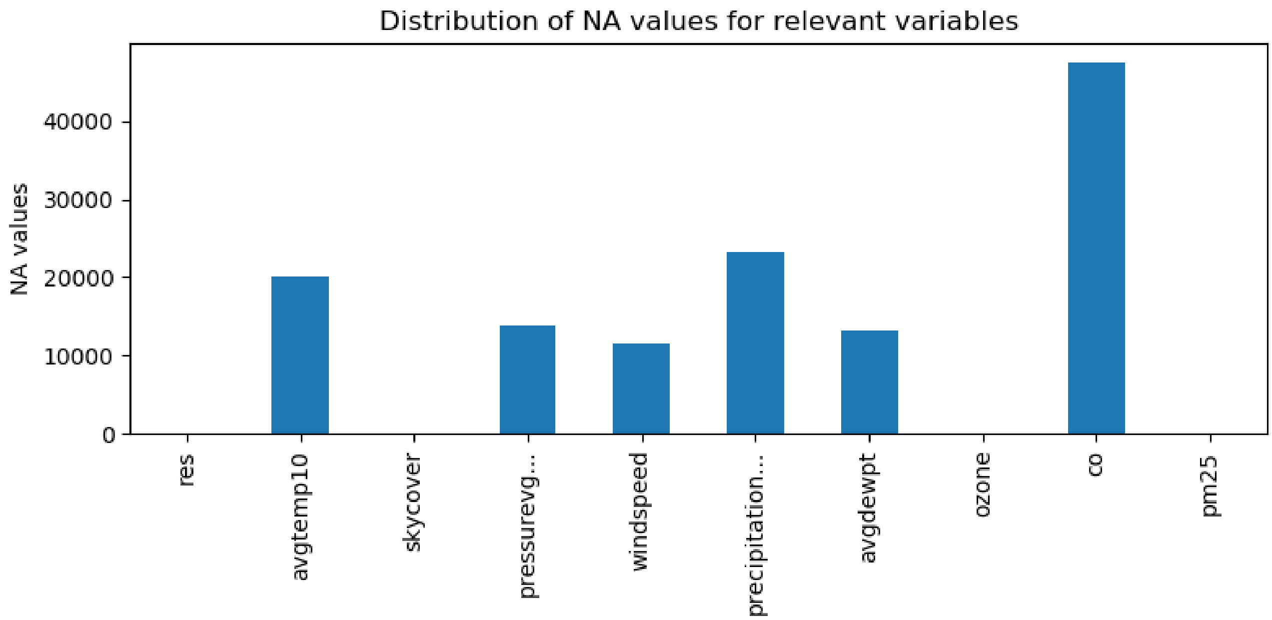
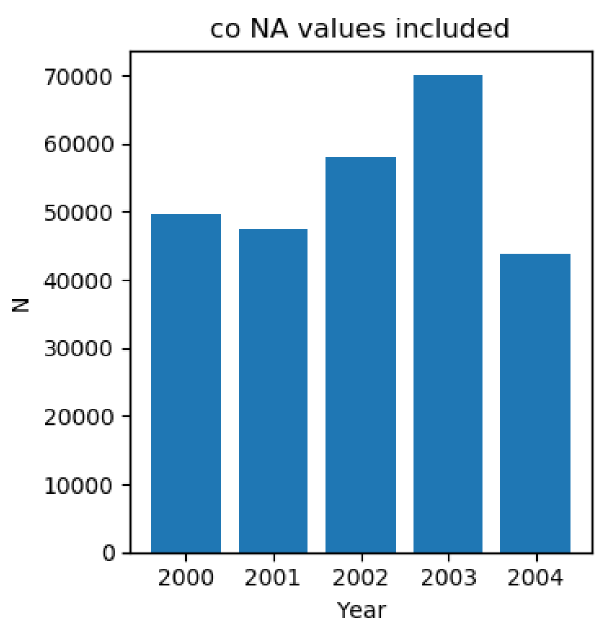
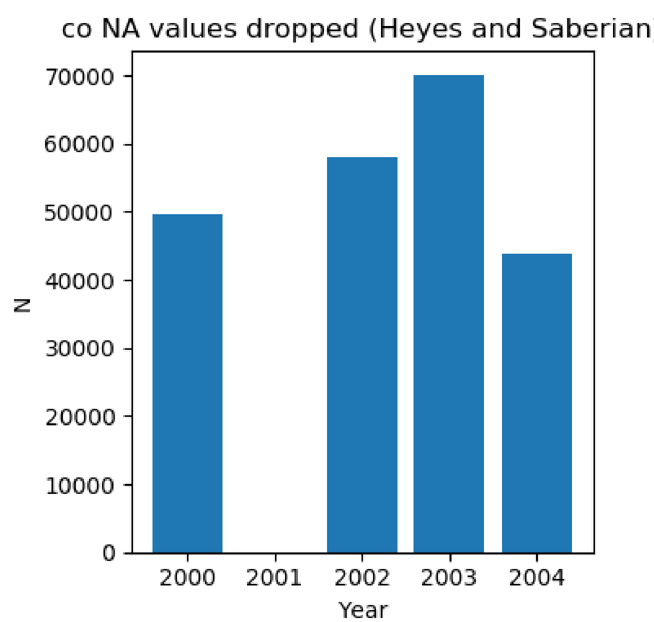


Figure B.2: Yearly observation count





C Authors response to referee's comments

I thank Vincent Loewe most thoroughly for his criticism and helpful suggestions on my paper. He points out that my yearly subsample approach may yield estimates with too high variances. It is true that it is somewhat problematic to prove that a coeficent that is not significantly different from 0 was subject to enough degrees of freedom in its estimation. I therefore did a test run of the yearly analysis with random subsamples (20k instead of 45k obs). This yielded similar results, which shall serve as somewhat of a proof that the yearly subsamples are sufficient to deliver meaningful results. As this was to be expected given the sample size, I do go into furhter detail in the paper itself.

In the discussion chapter I now include a short note on the issue that a linear probabilty model is prone to fitted values falling outside of the [0; 1] interval, as suggested. I explain that the LPM does not seem problematic in this analysis and its results are comparable to other papers, which apparently did not have problems with that, either.

Vincent explains that my example of judges' political sentiment is not adequate from a theoretical perspective to motivate the inclusion of latent factors. I agree, and changed the example to judges' general feeling of well-being and mood, which is more fitting. Apart from its motivation, my colleague generally questions the usefulness of the latent factor specification. As put in the discussion chapter, this additional method as well as its motivation serve only as an example how the given dataset could be further mined to maybe yield better estimates for temperature effects. Obviously, getting an alternative method such as the latent factor example to run properly would be beyond the scope of this paper. Even latent factors might not improve inference, assuming that in this very dataset considerable temperature effects truly exist. Nonetheless, the fact that after including latent factors significant temperature effects prevail, is a valuable finding in itself, even though it contradicts its motivation. I therefore find its inclusion as an example appropiate.

Another suggestion by Vincent is testing different dimension combinations for error clustering in the pooled regression. At an early stage of this paper I tested the estimation of the base model without any error clustering at all, which lead to similar results. That motivated me to omit potential issues of clustered errors and simply use them without any critical reflection. I now tested several different dimensions for clustering and still got similar results. Subsequently, I stay with my initial impression that this is not impactful enough to be further elaborated in the paper.

References

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Chen, Y., Tao, M., and Liu, W. (2020). High temperature impairs cognitive performance during a moderate intensity activity. *Building and Environment*, 186:107372.

Evans, S. and Siminski, P. (2021). The effect of outside temperature on criminal court sentencing decisions.

Gavresi, D., Litina, A., and Makridis, C. (2021). Split personalities? behavioral effects of temperature on financial decision-making. *SSRN Electronic Journal*.

Hadi, R. and Block, L. (2019). Warm hearts and cool heads: Uncomfortable temperature influences reliance on affect in decision-making. *Journal of the Association for Consumer Research*, 4(2):102–114.

Heyes, A. and Saberian, S. (2019). Temperature and decisions: Evidence from 207,000 court cases. *American Economic Journal: Applied Economics*, 11(2):238–265.

Holger Spamann (2020). No, judges are not influenced by outdoor temperature (or other weather): Comment.

Liu, H., Yang, J., and Yamada, Y. (2020). Heat and fraud: evaluating how room temperature influences fraud likelihood. *Cognitive Research: Principles and Implications*, 5(1).

Ryan, M. E. (2020). The heat: temperature, police behavior and the enforcement of law. *European Journal of Law and Economics*, 49(2):187–203.

Stevens, H. R., Graham, P. L., Beggs, P. J., and Hanigan, I. C. (2021). In cold weather we bark, but in hot weather we bite: Patterns in social media anger, aggressive behavior, and temperature. *Environment and Behavior*, 53(7):787–805.

Stroom, M., Kok, N., Strobel, M., and Eichholtz, P. (2021). Turning up the heat: The impact of indoor temperature on cognitive processes and the validity of self-report. *SSRN Electronic Journal*.

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

Zhang, F. and de Dear, R. (2017). University students' cognitive performance under temperature cycles induced by direct load control events. *Indoor air*, 27(1):78–93.

Zhang, F., Haddad, S., Nakisa, B., Rastgoo, M. N., Candido, C., Tjondronegoro, D., and de Dear, R. (2017). The effects of higher temperature setpoints during summer on office workers' cognitive load and thermal comfort. *Building and Environment*, 123:176–188.