Referee: 1

Comments to the Author

The paper makes a contribution just by showing how little correlation there is between various business climate indexes. If the indexes actually measured an objective characteristic called business climate, one would think there would be a strong correlation among various measures, but the paper points out there is not.

Thus, the results that these indexes have little ability to predict future economic growth is not surprising. Nonetheless, the paper's methodology of comparing across state borders makes a very good and very convincing test. The fact that eight different measures of growth are used shows the robustness of the paper's conclusions.

One explanation for these results is that there are state-specific factors other than business climate that affect economic performance. Another is that the indexes are poor measures of business climate. One might also be inclined to think that five years is too short a period to detect changes in business climate, but the fact that the indexes are very persistent over time seems to weigh against any argument that the time period chosen was too short.

The paper's conclusions would be very helpful to policy makers who consider using these indexes as a basis for public policy, in addition to their academic interest.

I don’t think there is anything to do here.

Referee: 2

Comments to the Author

This paper in some ways extends the JRS paper by Kolko et al. (KNM), in this case estimating the effects of business climate indexes at state borders, on the argument (discussed below) that this gives cleaner identification of the effects of the policies captured by the indexes. It ultimately reaches conclusions fairly similar to KNM (although you wouldn’t know that from the intro), but its contribution could be much sharper if it hewed closer to the original paper and isolated the differences attributable to the identification strategy. Without that, it is very hard to know what to make of the results from the combined set of papers (and there is recent one by Neumark and Muz from the NBER). Comments follow:

Major comments

1. The key contention in this paper is that we can better identify the effects of the policies captured by business climate indexes by looking at narrow areas along state borders (matched cross-border counties). The idea has some intuitive appeal, but there are some problems.

As the author notes, this strategy has been used in recent minimum wage work (Dube et al.). However, the author seems unaware of a budding follow up literature that raises questions about whether this actually is a better identification strategy (and follow up defense by the original authors). Whatever one makes of that debate, I think there is no question that it establishes that one wants to test whether the assumption that cross-border controls are really better is true. There are numerous comments in the paper asserting that they are, but that is not enough. Note, by the way, that the argument in Dube et al. is quite different. It is about contemporaneous or recent shocks that they argue are similar on two sides of the border.

TO DO: Read Dube et al paper and related citations – summarize the debate -- add to our paper in methods section if appropriate, or as a footnote.

There are many studies that estimate policy effects using cross-border designs. The author should discuss these, and what does or doesn’t make sense about it in the current context.

**We agree, there are a number of studies that estimate policy effects using cross-border designs in addition to Dube, Lester and Reich’s (2010) paper on the minimum wage. For example, Fox (1986), Holcombe and Lacombe (2004), Hoyt and Harder (2005) McKinnish (2005, 2007), Coomes and Hoyt (2008), Billings (2009), Dhar and Ross (2012), and Kahn and Mansur (2013).**

**Many of these only examine select regions or MSAs, while our study employs all cross-border pairs in the contiguous United States. A review of these papers finds that the most common argument for using the cross-border design is similar to ours; that is, that the border matching technique holds constant many factors that can vary geographically, such as climate and other spatial amenities, culture, proximity to markets, and local labor market conditions, without using dummy variables. This permits the focus to be on differences in state policies, such as taxes.**

**Several of these papers compare their cross-border results with cross-sectional models. For example, Holcombe and Lacombe (2004) undertake such a comparison and conclude that while the results are roughly in the same direction, the magnitudes of the effects are different.**

**In addition, only some of these papers construct the dependent variable(s) as ratios or differences across the border as we do (for example, Fox(1986, Holcombe and Lacombe (2004), . We agree that the alternative approach, using a border pair fixed effect, is potentially problematic in the presence of spillover effects.**

The discussion of why exactly cross-border counties provide a better experiment is vague. See pp. 4-5 and elsewhere. What exactly is controlled for by doing the cross-border comparisons, and under what assumptions? The author instead lists everything that could possibly contaminate the state-level regression (endogeneity, heterogeneity, etc.), and then asserts without much argument (and no evidence!) that the cross-border approach is more reliable.

TO DO: this seems related to the above – should be taken care of by previous comment.

The author also refers to the value of comparisons with the many states a state may border. But I don’t see any way that comes out in the analysis.

TO DO: Clarify why having multiple borders for a given county is good (variation I think, right?).

There are clearly reasons the cross-border controls may not be good. There can clearly be positive spillovers, which would attenuate any effects. Does northern NJ not benefit from proximity to NYC? It might be useful to think about two extremes. Do we want to estimate policy effects using cross-border regions, or separate islands? That depends. If there are important spillovers then islands would be ideal. If there are important unmeasured sources of heterogeneity, that are more similar for cross-border regions, then the author’s design makes more sense. But how do we establish this? This comes back to the issue of working much harder to validate the design.

TO DO: Again, may be taken care of by above? Need to at least address the issue of spillover effects

The author seems to have in mind the idea that state border regions “compete” with cross-border regions, but state interiors do not. What regions actually compete with each other, and in what industries? What markets are local and what markets are national, and do the results differ for the former (if there is competition across state borders)? This is all too vague.

TO DO: Think this is about the ‘discontinuity design’ – if we are going to find an effect, it is most likely to be at the border.

There is an important issue of what the border regions look like. Near NY, they are very dense with lots of economic activity. In CA, the border regions are essentially empty. One would think this could matter a lot for the interpretation of the results. This is an issue that has been addressed in other work using cross-border regions. A related point is that if a cross-border region is urban, then cities may adopt policies to compete with cross-border regions, in which case the state business climate measure may simply be less relevant at the border, attenuating effects.

TO DO: How much economic activity in our border counties? It is about one third of total US employment, income, population.

Can we find how it has been addressed in other work (yellow highlighted)?

2. One thing the paper emphasizes is whether indexes do a better job predicting future growth or past growth. The author takes it as a sign that an index is “bad” if it appears to predict past growth just as well. In this particular context, that may not be very sensible. First, as the author notes, the intertemporal correlation of these indexes is very high, so there may be little difference between the index at t predicting growth from t-5 to t and the index at t-5 predicting this growth. Second, my sense is that the indexes have a lot of inertia and don’t always use up to date policy parameters. Given that, it would not be at all surprising if the indexes seem to predict past growth. Also, didn’t KNM do a similar kind of test as a way of assessing policy endogeneity, although predicting the indexes with past growth? The author should clarify the differences. Does that fact that only the New Economy Index does more to look backward suggest that in general there is not a policy endogeneity problem?

TO DO: Compare our backcasting approach to KNM – describe how same/different.

To the point about intertemporal correlation – would it address this point if we only used the first date of the index? What do we lose if we do that?

Most are just bad at everything …..

3. The paper at some points seems like a direct extension of KNM, and at other seems quite distinct. I think especially because JRS published the original paper, it would be far more useful to readers for the author to hew very closely, for at least part of the paper, to what KNM did, but then substituting the cross-border design to see if it makes a difference. Otherwise it is very hard to tell what is going on. There are many other places, too, where there is not sufficient attention to the overlap. For example, the paper discusses the content of the indexes, but that was discussed in detail in that paper. At the same time, things are a bit muddled with respect to comparing the two papers. The last paragraph on p. 2 sounds like the paper finds no evidence that the indexes predict anything, whereas the conclusion sounds much more like the findings from KNM. But since sample periods, regressions, and some indexes (including names of the same indexes) are not consistent, it is very hard to compare.

TO DO: More clearly describe how our paper complements and extends KNM. Check names of indexes for consistency with KNM, maybe add footnotes if different to help readers compare.

4. When the author discusses the fairly low contribution of the indexes to the R square, I don’t know what to make of this. Compared to what? For example, the regressions appear to have no other controls (which is weird in light of KNM), and we wouldn’t expect them to be important relative to the business cycle (there are no time effects). There aren’t many papers that make much of explanatory power. What is the rationale for doing so here?

TO DO: Need to clarify that we are interpreting the R2 as how much variation in the outcomes is accounted for my the business climax indices. We aren’t trying to explain growth in income for example, we are trying to see how well the indices predict the growth in outcomes.

And simply throwing them all in to establish an upper bound for explanatory power makes no sense at all. I can’t think of a statistical or conceptual rationale for this. Moreover, pointing to low R-squares to say the indexes are “just not that good” doesn’t really make sense. Again, what is the context and comparison?

TO DO: Better explain rationale here – think is that that maybe the indices are explaining different things and if we combine them, we might get a better prediction?

5. On p. 8 the author suggests that somehow the state border design helps with the problem that states may react to their neighbors’ policies. How does the state border design help on this score? Doesn’t it make the endogeneity worse?

TO DO: Check this and clarify.

6. Most empirical research on county-level data uses QCEW employment, and not much else. How good are the BEA regional data used in this paper? What are the data sources?

**The Bureau of Economic Analysis regional data are based primarily on administrative records data such as state unemployment insurance programs and state and federal income tax codes. The agency notes, “Using administrative records data and census data to measure local area personal income has both advantages and disadvantages. By using these data, BEA can prepare detailed annual and quarterly estimates at a relatively low cost and without increasing the reporting burden on businesses and households. However, because the source data often do not precisely match the concept being estimated, they must be adjusted to compensate for differences in definitions, coverage, timing, and geographic detail.”**

**http://www.bea.gov/regional/pdf/lapi2013.pdf**

**This page answers the question, what is the difference between BEA employment and wages and BLS and Census employment and wages?**

[**http://www.bea.gov/faq/index.cfm?faq\_id=104**](http://www.bea.gov/faq/index.cfm?faq_id=104)

**The key point is that BEA’s estimates of total employment and total wage and salary disbursement are derived from the BLS data. BEA makes adjustment to account for employment and wages not covered, or not fully covered, by unemployment insurance programs (for example, nonprofits organizations, students, elected officials, interns, farms etc.).**

7. The author notes that for some outcomes (like wages) there is no prediction that policies will create some differences at the border. Then why use these in the analysis?

TO DO: Respond to this comment – sort of a falsification test?

Minor comments

1. What does the author have in mind as state specific productivity shocks (p. 1)?

TO DO: Provide an example of a state specific productivity shock.

2. There needs to be more careful discussion of what we are estimating when we put the BC indexes on the right hand side of these kinds of regressions. KNM have some nice discussion of this, whereas it is left a little vague in this paper. This is another place where I can’t see any disadvantage to sticking much closer to the original paper and highlighting key differences in empirical strategy.

TO DO: Provide an explanation of what we are estimating – cite KNM – pgs 221 -222 (I think this is what is being referenced – idea is that indexes measure a bundle of policies, not specific policies)

3. What is the value of the many quotes arguing that the indexes compete? I don’t think the interesting question is simply which index is better.

TO DO: ??? We could remove them. I think the idea was to motivate why this is interesting?

4. On p. 5, the author says that indexes that included outcomes were not used in the analysis. But KNM showed that you can get the raw data and just strip these out. The same issue applies to those reported as rankings. Why not use the raw data as they did?

TO DO: Response here? I think one response is that the indices that include outcomes are confounded – point is not to make them better ( or more valid) but to evaluate performance of the ones that might reasonably measure policy levers that states can change that are expected to impact growth.

How the weights of the different indexes match what’s out there.

5. P. 6 says that all the indexes consider various aspects of quality of life. I disagree. Some narrowly focus on taxes and costs.

TO DO: Response????

6. Figure 1 was not included, and perhaps because of this I have no idea what the paragraph following the reference to the figure (p. 10) means. And why is the comparison to clustering of interest? Do these two approaches in any way address the same problem? (Of course clustering doesn’t affect the point estimates anyway.)

TO DO: Check what figure 1 is/was and either include or revise as needed.

Reply how clustering and weighting address the same problem – suppose we could make this discussion a footnote?

7. P. 18 states that the indexes lack “the scientific rigor typically required of social science research.” What does this mean, exactly? How can one possible assess rigor based on whether or not we get predictive power? Maybe the theory is wrong.

TO DO: either rewrite, or respond to this.

8. I have no idea what the author’s assertion about indexes performing best immediately after issuance (p. 15) means.

TO DO: clarify this – I think the idea is that over longer periods of time, unforeseen changes in technology or other factors make prediction more difficult? The farther out you go, the more likely you are to miss?

9. P. 16 describes the effect of a 10% increase in a BC index as increasing growth by less than 1% as small. Is a 1% effect small, if this is a growth rate? That seems large to me but I may be misinterpreting the magnitude.

TO DO: Revise to make the interpretation more precise?

10. I don’t see any basis in the paper for the conclusion in the last sentence. What does the author have in mind regarding forecasting tools? Using the indexes differently? If so, then why not do that?

TO DO: Revise the last sentence to be more specific about what is meant here.

Referee: 3

Comments to the Author

Referee report on paper on business climate indices

Summary Comments

This is a useful paper that shows that business climate indices largely do not provide much new information. I view one revision as essential, others as optional.

Specific Comments

1. The paper should mention more prominently that the elasticity of business activity at state borders with respect to “business climate” (say business cost measures) does NOT measure the elasticity of business activity overall for the state. The elasticity at state borders shows the response of businesses holding variables constant such as the local labor market’s wages and productivity. But the elasticity overall with respect to state business climate should not hold that constant. Overall for the state, there are likely to be more businesses that are inframarginal, and whose location decisions will not be affected by business cost factors. What this paper is exploring is whether state “business climate factors” will affect a state’s growth if we could imagine the state being identical otherwise to all other states, which is not the case. What is measured here is closer to the “within metropolitan area” elasticity of business activity with respect to business costs, which is likely to be much higher than the “across metropolitan area” elasticity of business costs. This should be emphasized in a revision.

TO DO: Clarify language – I think this one will happen with the revisions requested by Referee 2 above w.r.t. the strengths of the border design and what it is we are actually measuring.

2. As a more optional revision, I was surprised that for all these regressions, there was not also exploration of whether these indices predicted future change in variable y holding constant past change in variable y. This seems like a more intuitive test of the proposition of whether these indices add anything that we don’t already know from looking at past trends.

TO DO: Should we try this? If so, Kevin should re-estimate these – maybe do one set as a robustness check?

3. As another more optional revision, I was surprised that there wasn’t exploration of whether the change over time in these indices was related to changes over time in trends in the various dependent variables. If these indices really are measuring things, then changes over time in these indices should be related to changes over time in the growth of the various dependent variables. Alternatively, the dependent variables and the indices could be pooled and state fixed effects added.

TO DO: Respond to this: I think the response here is that we are trying to isolate the predictive power of the index, not explain/predict the actual growth in the outcomes.