EPI

ISSUE BRIEF

ECONOMIC POLICY INSTITUTE | ISSUE BRIEF #308 JULY 19, 2011

FLAWS CALL FOR REJECTING CRAIN AND CRAIN MODEL

Cited $1.75 trillion cost of regulations is not worth repeating

BY JOHN IRONS AND ANDREW GREEN

Executive summary

Among the talking points used by critics of regulations is the alleged $1.75 trillion cost of regulations in 2008. This estimate comes from a study written by Nicole V. Crain and W. Mark Crain for the Small Business Administration's Office of Advocacy. A substantial majority of these costs—$1.2 trillion or 70 percent—are based on the author's use of an econometric regression analysis to determine the costs of "economic" regulations, such as those rules affecting the financial industry.

This paper examines the econometric regression analysis that Crain and Crain used to determine the costs of economic regulations. We conclude that the regression model is so conceptually flawed and statistically fragile that its findings should be rejected.

Flawed methodology

More specifically, the Crain and Crain methodology for determining the cost of economic regulations contains the following fundamental errors:

- It fails to capture the timing between changes in the determinants of economic growth and the amount of ensuing growth. The estimation does not properly take into account time-series dynamics.
- It misses the potential reverse causality between factors associated with economic growth and the growth itself. For example, the Crain and Crain model considers only how such factors as education, the extent of broadband use, and regulations affect economic growth without considering how economic growth can affect education, broadband use, and regulations.

- It uses a composite World Bank measure of "regulatory quality" that may capture a range of factors that could lead to higher levels of economic activity but have nothing to do with the stringency of regulations in a country. Indeed, one of the authors of the World Bank's "Index of Regulatory Quality" disputes the way it is used in the Crain and Crain study.

As broad evidence of its conceptual flaws, the Crain and Crain regression analysis finds that a country's economy shrinks as the level of education of its population grows. To unquestioningly accept the finding that economic regulations cost $1.2 trillion, one must also believe that more education somehow undermines economic growth.

Flawed data

The results of the Crain and Crain study, even aside from flawed methodology, appear to be a result of a flawed data set. In particular:

- The Crain and Crain data set is missing close to half of the potential observations. The study purports to use data describing various indicators in Organization for Economic Cooperation and Development countries from 2002–2008 in order to determine the relationship between regulation and GDP. This implies a potential total of 210 (seven years times 30 countries) "observations." Yet so much information is missing that 92 (44%) of the observations had to be dropped from the regression model.
- Missing data is primarily due to the education measure used, namely primary school completion rates.
- Among the observations dropped are all observations from five countries and one entire year (2008). Countries such as Austria are retained, but only partially represented due to incomplete data. For example, while Austria's information for 2002, 2004, and 2007 are used, information for 2003, 2005, 2006, and 2008 had to be dropped out of the data set.

Econometric regressions of this type relying on time-series or panel data sets with large numbers of missing observations are prone to yielding peculiar results. In short, one cannot confidently describe a relationship between regulation and GDP when the countries and years used to determine that relationship are only sporadically represented.

As broad evidence of its statistical fragility, an improved application of Crain and Crain's still-flawed conceptual method yields a far different finding from Crain and Crain's application. We update the Crain and Crain study with data for 2008 and fill in nearly all of the missing data points for earlier years by generating a more complete education variable.

We find that in this more complete data set there is no statistically significant relationship between regulatory quality and GDP, meaning that even Crain and Crain's own flawed model does not provide reliable evidence of an impact on economic activity. (While this use of updated and more complete information yields sounder estimates than the Crain and Crain analysis, it still contains the other conceptual flaws with the model, and still yields the unsupportable finding that more education leads to smaller economies.)

Because the Crain and Crain results are driven by a combination of poor data and a flawed empirical approach, the report should not be used either as a valid measure of the economic costs of regulation or as a guide for policy.

Introduction

One of the most widely used studies purporting to show extraordinarily large economic costs of regulation was prepared by Nicole V. Crain and W. Mark Crain (Crain and Crain 2010) for the Small Business Administration's Office of Advocacy. Many policymakers cite one of the study findings that federal regulations cost more than $1.75 trillion in 2008. The large majority (70%) of the cost is based on a regression analysis that sought to determine the costs of "economic" regulations, such as financial regulations.

The overall approach employed by Crain and Crain contains a series of conceptual and empirical problems as identified by the Center for Progressive Reform (Shapiro, Ruttenberg, and Goodwin 2011), the Congressional Research Service (2011), and the Economic Policy Institute (Shapiro and Irons 2011).[1] The Obama administration recently disavowed the Crain and Crain study, stating that it "wildly overstates the cost of regulation" and has "very serious methodological problems and is out of step with mainstream economists" (Obama administration 2011). The administration's Council of Economic Advisers found the $1.75 trillion figure to be "utterly erroneous" (Goolsbee, 2011).

This issue brief examines the conceptual and empirical problems of the Crain and Crain regression analysis and the bottom-line results which make up the bulk of the costs identified by the authors.[2]

Problems with methodology

Crain and Crain (equation 1) summarizes the relationship between GDP per capita and so-called regulatory quality. Specifically, equation 1 is the econometric specification in which GDP per capita is explained by the World Bank's "Index of Regulatory Quality" and other variables.[3]

$$(1) \quad \text{GDP per Capita}_{it} = \beta\,(\text{World Bank Index of Regulatory Quality})_{it} + \phi(X)_{it} + \alpha_i + \varepsilon_{it}$$

The regulatory quality index ranges in value from -2.5 to 2.5 and is based on a combination of surveys and expert-based judgment measures (Kaufmann, Kraay, and Mastruzzi 2010).

The regulatory index is meant to capture "perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development."[4] The index includes sub-indexes from other organizations, including from *The Economist* magazine and the Heritage Foundation, as well as indexes derived from surveys.[5] Questions have been raised about the meaning and interpretation of the index, for example, CRS (2011) notes that "one of the authors of the regulatory quality index said that Crain and Crain misinterpreted and misused the index, resulting in an erroneous and overstated cost estimate." Of interest, Denmark, Finland, Netherlands, and Sweden—countries not typically associated with a lack of regulation—all scored higher on the index of regulatory quality than did the United States over the seven-year period; that is, this index of regulatory quality is not necessarily a fully accurate measure of regulatory stringency (though Crain and Crain use it as such).

The core Crain and Crain finding on economic costs come from a positive, statistically significant coefficient on the regulatory quality index (see Table 3 for full estimation results); this is interpreted as an indication that the higher a country's score on the regulatory quality index, then the higher its GDP per capita.

The econometric specification includes additional explanatory variables (X) that would determine economic growth. In the Crain and Crain specification, these consist of the (natural log of) country population, school completion rates as a share of the eligible population,[6] fixed broadband subscribers per 100 people, and foreign trade as a share of GDP. The regression also contains country fixed effects, and a dummy time variable for 2007.

TABLE 1

OECD countries included in Crain and Crain data and regression

Included in Crain and Crain regression model (25)

Austria	Finland	Ireland	Netherlands	Spain
Belgium	Germany	Italy	Norway	Sweden
Canada	Greece	Korea, Rep.	Poland	Switzerland
Czech Republic	Hungary	Luxembourg	Portugal	Turkey
Denmark	Iceland	Mexico	Slovak Republic	United States

Included in inital Crain and Crain data set, but dropped from regression model (5)

Australia	France	Japan	New Zealand	United Kingdom

OECD countries not in data set or regression model (4)

Chile	Estonia	Israel	Slovenia

SOURCE: Author's analysis of Crain and Crain (2010), and OECD.

Sample

Crain and Crain indicate that they examine data from OECD countries from 2002 to 2008. As **Table 1** indicates, of the 34 countries that are now part of the OECD, four were not OECD members at the time of the study, and five were dropped because they were missing data values in each of their seven yearly observations. The five countries dropped from the regression analysis due to data limitations are Australia, France, Japan, New Zealand, and the United Kingdom.

Possible misspecification

The specification in equation 1 raises several conceptual and data issues. First, the specification lacks a realistic specification of time-series dynamics. For example, this specification assumes a *contemporaneous* causal link between school completion rates or regulation and GDP per capita. Lagged variables or start-of-period measures are not considered. In contrast, standard models of cross-country growth allow for the time that it takes for various factors (such as education, investment, etc.) to impact the economy, by, for example, relating economic *growth* to the value of the determinants at the start of the period under consideration (Barro 1991). (Also, because the GDP measure is likely to be non-stationary, the specification in equation 1 is likely to pick up spurious correlations with other time-trending variables.)

It is also important to remember that because the model contains country fixed effects, the relation between education and GDP is determined by within-country variation over time for only six years, meaning that it is essential to properly specify the time-series dynamics in order to get meaningful results.[7]

This overall dynamic misspecification might be responsible for the fact that Crain and Crain find a *negative* coefficient on their measure of education. According to their analysis, higher levels of primary school completion lead to *lower* levels of GDP in that same year; a finding that stands in stark contrast to the established literature on economic growth.

Second, the specification in equation 1 mischaracterizes as one-way the causal direction of GDP and the explanatory variables (or "covariates"). For example, the specification implicitly assumes that broadband penetration causes higher levels of GDP, which could be true, but it also assumes that higher GDP is not a causal factor in determining broadband use, a highly dubious assumption. The same is true for other variables, if GDP levels in part cause differences in regulation, education, trade, etc., then the Crain and Crain specification would not yield valid results.

Third, the regulatory index could be correlated with a variety of other factors that might impact GDP, such as environmental factors, federal investment policies, workforce policies, etc. As such, the regulatory measure in this analysis would capture a variety of impacts outside of regulatory policy per se.

It is unclear how to interpret the core result in light of the likely dynamic model misspecification, causality issues, and the vagueness of the regulatory index discussed above. Compounding this uncertainty, the particular Crain and Crain findings result from a data sample that is missing a large number of data points.

Problems with the data

At noted earlier, Crain and Crain's initial data set included 30 countries and spanned seven years, yielding a total potential sample of 210 observations. Of these potential country-year observations, 92 observations (44%) were missing at least one piece of data and were therefore dropped from the data sample for the regression analysis, leaving just 118 observations in the Crain and Crain estimation. As a result, five countries (Australia, France, Japan, New Zealand, and the United Kingdom) and one entire year (2008) were dropped from the sample, producing a sample that is a patchwork of country-years (see **Table 2**).

A closer look at the data reveals that the choice of education variable, primary school *completion* rate, is to blame for most of the missing data. Additional missing values arise from missing country-years in the trade/GDP measure.

This patchwork of observations is disconcerting, in part, because the empirical specification uses the within-country variation to identify the impact of the regulatory index. It is also troubling that an entire year, 2008, is missing since there was a very small sample to begin with.

To correct for this problem, we recalculated equation 1 with two changes. First, we updated the trade variable as a share of GDP data from the same source as Crain and Crain (World Bank, World Development Indicators, online database), which allowed us to add in data from 2008.

Second, we generated a more complete school-completion rate data set. Using data on school enrollment, we estimated equation 2 (with country fixed effects) to predict school completion for the missing observations.

$$(2) \quad \ln(\text{school completion})_{it} = \beta\,(\ln \text{school enrollment})_{it} + t + \alpha_i + \varepsilon_{it}$$

As in the original Crain and Crain regression, we include country fixed effects. We also include a linear time trend to better fit the data.[8] Using the more complete data set (205 total observations out of a possible 210), we re-estimated Crain and Crain's equation 1. The results are shown in **Table 3**. The results are qualitatively similar, except that the coefficient on the regulatory index is statistically insignificant at the 5% level, with a substantially smaller point estimate.

The results strongly suggest that the finding of a significant impact of the regulation quality index on GDP is driven by the particular pattern of missing data in the initial data set: The same analysis with a more complete data set yields no significant impact of the regulatory index on GDP. And while using the more complete data set yields empirically more reliable results, it does not correct for the other methodological flaws with the Crain and Crain approach, underscored by the fact that even when the more complete data set is used, more education is still associated with smaller economies.

TABLE 2

Matrix of all possible (210) observations in regression model

If 1, at least one variable has a missing value; else 0

Country	2002	2003	2004	2005	2006	2007	2008	Total dropped observations
Australia	1	1	1	1	1	1	1	7
Austria	0	1	0	1	1	0	1	4
Belgium	1	1	1	1	0	0	1	5
Canada	1	1	1	1	0	1	1	6
Czech Republic	0	0	0	0	0	0	1	1
Denmark	0	1	0	0	0	0	1	2
Finland	0	0	0	0	0	0	1	1
France	1	1	1	1	1	1	1	7
Germany	0	0	0	0	0	0	1	1
Greece	1	0	0	0	0	0	1	2
Hungary	0	0	0	0	0	0	1	1
Iceland	0	0	1	0	0	0	1	2
Ireland	0	0	0	0	0	1	1	2
Italy	0	0	0	0	0	0	1	1
Japan	1	1	1	1	1	1	1	7
Korea, Rep.	0	0	0	0	0	0	1	1
Luxembourg	0	0	0	0	0	0	1	1
Mexico	0	0	0	0	0	0	1	1
Netherlands	0	0	1	1	1	1	1	5
New Zealand	1	1	1	1	1	1	1	7
Norway	0	0	0	0	0	0	1	1
Poland	0	0	0	0	0	0	1	1
Portugal	1	1	0	1	1	1	1	6
Slovak Republic	0	0	0	0	0	0	1	1
Spain	1	1	0	0	0	1	1	4
Sweden	0	0	0	0	0	0	1	1
Switzerland	0	0	0	0	0	1	1	2
Turkey	1	1	0	0	0	0	1	3
United Kingdom	1	1	1	1	1	1	1	7
United States	0	0	0	0	0	1	1	2
Total	11	12	9	10	8	12	30	92

SOURCE: Author's analysis of Crain and Crain (2010).

TABLE 3

Estimation of Equation 1 with Crain and Crain sample data, and more complete data

	Dependant variable: Ln (GDP per capita)	
	Crain and Crain	*Augmented data*
Number of observations/total possible	118/210	205/210
Number of countries	25	30
Independent variable		
Regulatory Quality Index	0.094*	0.036
	(0.034)	(0.021)
Ln(population)	0.085	-0.060
	(0.228)	(0.144)
Ln (trade/GDP)	0.241*	
	(0.049)	
Ln (trade/GDP) updated		0.194*
		(0.029)
Ln (primary school completion rate)	-0.282*	
	(0.098)	
Ln (primary school completion rate) Actual+predicted		-0.161*
		(0.070)
Ln (fixed broadband / 100 people)	0.033*	0.039*
	(0.004)	(0.003)
Constant	8.39*	10.59*
	(3.80)	(2.41)
R^2 Within	0.855	0.814
R^2 Between	0.031	0.088
R^2 Overall	0.015	0.089

* Significant at 5% level. Number in parentheses are standard errors. Each model includes country fixed effects and a dummy for 2007.

SOURCE: Author's analysis.

Conclusion

These findings suggest that the original Crain and Crain results are driven by a combination of poor data, and a flawed empirical approach. In short, Crain and Crain found that economic regulations cost $1.2 trillion in 2008 because missing data in the initial data set and a misspecification of the relationship between the variables led to a spurious correlation between their chosen measure of regulatory quality and GDP. As such, the report's headline $1.75 trillion estimate should not be used either as a valid measure of the costs of regulation or as a guide for policy.

Endnotes

1. These and other analysts had reported problems replicating the precise data analysis used to generate the broader economic costs (specifically the regression as reported in Table 2, page 23 of the Crain and Crain report). Subsequently, EPI was able to obtain the original data used in the analysis and to closely match the regression results (see the appendix).
2. Other aspects of the paper have been explored elsewhere, including in Shapiro, Ruttenberg, and Goodwin (2011).
3. Further details on replicating the precise estimates in Crain and Crain are available in Irons and Green (2011) "Memorandum re: Crain and Crain (2010) Results Replication" available at http://www.epi.org/page/-/Crain%20and%20CrainMemo_FINAL.docx
4. See information at http://info.worldbank.org/governance/wgi/pdf/rq.pdf
5. See CRS (2011) for a fuller discussion.
6. The Crain and Crain paper cites only school enrollment as a variable; however, their data set as obtained by the authors contains school completion data. In a response to our earlier analysis, Crain and Crain claim this is just a typo "in one place in the text." However, school "completion" was never cited anywhere else in the paper.
7. The sample is only six years, not seven, because there is no data on school completion in 2008, effectively dropping that year from the sample.
8. The estimated equation is Ln (completion) = 1.46 + 0.68 (ln enrollment) - 0.0048t + α_i , with all coefficients statistically significant at the 5% level or better.

References

Barro, Robert. 1991. "Economic Growth in a Cross Section of Countries." *The Quarterly Journal of Economics.* Vol. 106, No. 2, pp. 407-443.

Congressional Research Service (CRS). 2011. *Analysis of an Estimate of the Total Costs of Federal Regulations.* Washington D.C.: CRS; http://www.progressivereform.org/articles/CRS_Crain_and_Crain.pdf

Crain, Nicole V. and W. Mark Crain. 2010. *The Impact of Regulatory Costs on Small Firms.* Washington, D.C.: Small Business Administration Office of Advocacy; http://archive.sba.gov/advo/research/rs371tot.pdf

Goolsbee, Austan. 2011. "A 21st Century Regulatory System." *The White House Blog.* June 23.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010. *The Worldwide Governance Indicators: Methodology and Analytical Issues.* World Bank Policy Research, Working Paper No. 5430. Washington, D.C.: World Bank; http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1682130

Shapiro, Isaac and John S. Irons. 2011. *Regulation, employment, and the economy: Fears of job loss are overblown.* Economic Policy Institute, Briefing Paper #305. Washington, D.C.: EPI; http://www.epi.org/publications/entry/regulation_employment_and_the_economy_fears_of_job_loss_are_overblown

Shapiro, Sidney A., Ruth Ruttenberg, and James Goodwin. 2011. *Setting the Record Straight: The Crain and Crain Report on Regulatory Costs.* Center for Progressive Reform, White Paper #1103; February 2011, at http://www.progressivereform.org/articles/SBA_Regulatory_Costs_Analysis_1103.pdf

Obama administration. 2011. "Comments on Regulations and Snowe Amendment No. 390." Unpublished executive branch document, June 6.