The Commissar for Traffic Presents the Latest Five-Year Plan

Andrew Gelman and Phillip N. Price

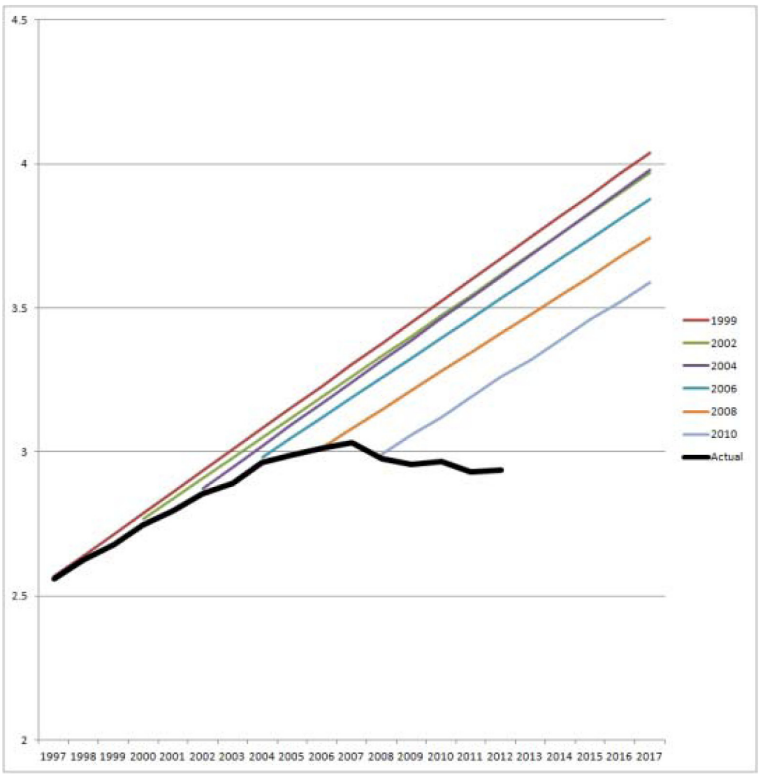


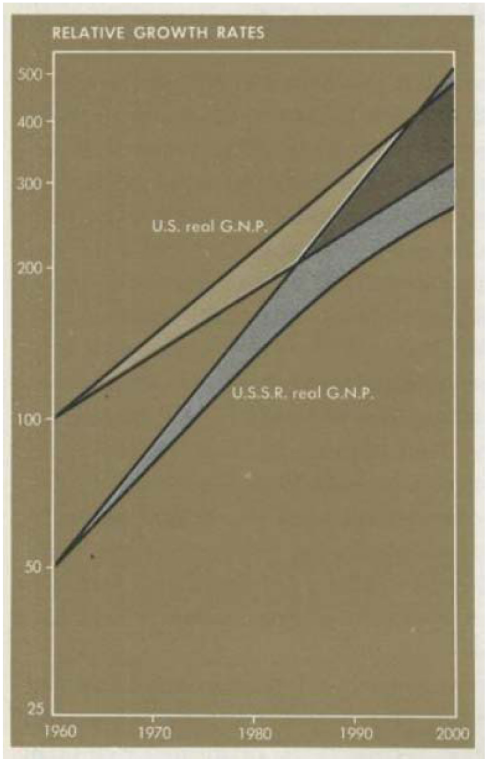
Figure 1. The same forecast repeated year after year, even after being falsified by data

Figure 1, taken from the State Smart Transportation Initiative, shows, on the y-axis, trillions of vehicle-miles traveled in the United States for each year since 1997, as well as 20-year traffic projections made approximately every two years by the U.S. Department of Transportation (USDOT). Each of the rising colored lines represents a forecast made in a different year. The black line represents actual traffic trends on U.S. roads, which, over this period, never rose as quickly as the forecasters had predicted and actually started a modest decline in 2007.

As Clark Williams-Derry of *Sightline Daily* pointed out in a discussion of this plot:

> This is not entirely USDOT's fault, though. These forecasts are a "roll-up" of forecasts made by state DOTs. The U.S. agency just collects the forecasts and reports them to the public: garbage in, garbage out.
>
> But in a way, that's even more sobering than if the fault were localized in USDOT, since it provides clear and compelling evidence that the nation's entire transportation forecasting apparatus is completely broken. In the aggregate, all of those hard-working forecasters in all those state DOTs are just making up numbers.

Indeed. We're reminded of this notorious graph from the legendary Paul Samuelson economics textbook (from 1961)



In contrast to the predictions shown in the graph, real GNP in the USSR remained at roughly half the U.S. level until the USSR dissolved in 1991[1].

Of course, there's nothing unethical or even embarrassing about making a wrong prediction. But as Alex Tabarrok points out, Samuelson revised his economic predictions year after year in the same way the Department of Transportation has revised their traffic predictions:

> In subsequent editions, Samuelson presented the same analysis repeatedly, except the overtaking time was always pushed further into the future so, by 1980, the dates were 2002 to 2012. In subsequent editions, Samuelson provided no acknowledgment of his past failure to predict and little commentary beyond remarks about "bad weather" in the Soviet Union.

As Tabarrok and his commentators note, this mistake can't simply be attributed to socialist sympathies of the center-left Samuelson: "For one thing, various other leftist economists did not think the Soviets were catching up, and another thing, political commentators on the right at the time were all telling us that the communists were about to overwhelm us militarily."

Aiming for a Good Particular Forecast vs. Using a Conventional Approach

Why are we discussing these problematic projections in a column about ethics? These projections are incorrect and maybe even ridiculous, but where does ethics come into it?

Let's start with the fact that projections matter. It's impossible to make perfect projections and, in some circumstances, it may be impossible to make good projections, but it is important to try to make good projections. As Eric Sundquist of SSTI points out with regard to the traffic forecasts:

> Not only are these data being aggregated into the national report, but they also are being used in project selection and development around the country. High estimates of VMT have several negative implications, including 1) they imply a level of "needed" spending that is politically unachievable, 2) they can spur overbuilding on projects, draining resources from critical preservation and multimodal investments, and conversely 3) they can discourage construction of lower-cost, lower-throughput streets that improve livability and property values.

Are the Department of Transportation employees trying to make good projections? We don't know for sure, but we don't think so. The projections aren't just wrong in the same direction, they're almost all the same: They all show projected growth beginning in the year the projection was made and at almost exactly the same number of miles per year. The projections seem to have been created by taking the historical growth rate in traffic miles from the 1980s and early 1990s and extrapolating it into the future, an approach that worked fine over the previous 10 years but stopped working around 1997.

The fact that the 1999 and 2002 projections ended up being much higher than reality didn't necessarily indicate something was wrong with the approach the modelers were taking: two- or three-year plateaus in traffic miles had happened before, but the upward march of mileage soon resumed. But by 2006, the modelers should have been asking themselves what was going wrong, and the 2008 and 2010 projections seem unconscionable. There's no theoretical reason to believe traffic volumes should always increase linearly, so once this empirical rule stopped working, it should have been abandoned.

Similarly, Samuelson's initial forecasts may well have made sense at the time he made them. But after multiple years of the Soviet economy not reaching its projected growth rate, it would have been appropriate for him to

update his model. Indeed, this admission of error and reassessment could well have been an instructive addition to his textbook.

This connects to a more general distinction between two goals of statistics, which might be labeled "get the right answer now" and "use an accepted approach."

There is always a need for balance between these goals. Consider, for example, the study of public opinion. The conventional approach relies on formal random sampling protocols and predetermined data analysis methods based on the statistical literature. But the right answer now can require deviating from accepted protocols. Sticking with the conventional approach can be unworkable or lead to clearly wrong answers—for example, the practical difficulties of nonresponse imply we will never get a pure random sample, so we may need to improvise in our analysis—but too much freedom in data analysis can lead to the sort of opportunistic adjustment that would make it difficult to ever discover something surprising or new.

For another example, consider the decision of how to handle undercount in the U.S. Census. If the goal is to get closer to the right answer, you will want to adjust the raw counts. But this is controversial in part because it involves deviating from the standard of pure enumeration: There is a concern that too much flexibility can lead to serious error and perhaps a preference for "the devil you know."

What was going on with the Commissar for Traffic and the economics textbook writer in the ridiculous forecasts discussed above? It is possible they repeated their wrong linear forecasts over several years out of simple laziness, or, worse, out of a considered decision to inflate the forecasts for budgetary or ideological reasons. But a more charitable explanation is that they felt they were following the method that had been accepted and worked fine in previous years. From this perspective, once you have a linear forecast, you have to keep going with it until there is overwhelming evidence against it. Don't let the data get in the way of a good theory and all that.

Another rationale for using a linear forecast is its simplicity. Even if the predictions from such a model are clearly biased, users might prefer a forecast they can understand. Richard Smith, in a New England Symposium on Statistics in Sports talk, illustrates with an example in which he was asked by the organizers of the Boston Marathon to project finishing times from runners who did not complete the race because of the bombings. He constructed a sophisticated prediction algorithm, but the marathon organizers ended up going with a simple linear projection, even though it performed quite a bit worse under cross-validation, possibly because the simpler approach was easier to explain and defend. We have seen similar choices made, for example, in value-added assessment in education, where policymakers feel more complicated ratings and decision rules will be more vulnerable to criticism.

Nonetheless, we think the Department of Transportation in recent years, and Paul Samuelson in the 1960s, made the wrong decision to hold on too long to clearly biased predictions in settings in which forecasts have real consequences.

Getting the wrong answer isn't unethical. Having a statistical or physical model that makes bad predictions isn't unethical. What seems unethical to us, both in the case of the textbook writer in the 1960s and the commissars today, is not recognizing the problem, even after the forecasting method has been completely destroyed by the data.

[1] This fact surprised us. Wasn't the Soviet economy supposed to have "collapsed"? In fact, as measured by Gross Domestic Product, the Soviet economy grew at a slightly higher rate than that of the United States from 1960–1980 and at a slightly lower rate from 1980–1990.

Further Reading

Levy, D. M., and S. J. Peart. 2011. Soviet growth and American textbooks: An endogenous past. *Journal of Economic Behavior and Organization* 78:110–125.

Samuelson, P. A. 1948–1980. *Economics: An introductory analysis.* New York: McGraw-Hill.

Sundquist, E. 2013. New travel demand projections are due from U.S. DOT. Will they be accurate this time? State Smart Transportation Initiative, Dec. 16, *www.ssti.us/2013/12/new-travel-demand-projections-are-due-from-u-s-dot-will-they-be-accurate-this-time*.

Smith, R. 2013. New England Symposium on Statistics in Sports—Talk by Richard Smith. Video at *www.youtube.com/watch?v=cRUfrkbXUNQ&feature=youtu.be*.

Tabarrok, A. 2010. Soviet growth and American textbooks. Marginal Revolution blog post, Jan. 4. *http://marginalrevolution.com/marginalrevolution/2010/01/soviet-growth-american-textbooks.html*.

Williams-Derry, C. 2013. Traffic forecast follies: The U.S. DOT refuses to learn from recent travel trends. Sightline Daily blog post, Dec. 23. *http://daily.sightline.org/2013/12/23/traffic-forecast-follies*.

About the Author

Andrew Gelman is a professor of statistics and political science and director of the Applied Statistics Center at Columbia University. He has received many awards, including the Outstanding Statistical Application Award from the American Statistical Association and the award for best article published in the *American Political Science Review*. He has coauthored many books; his most recent is *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*.