A Brief Parable of Over-Differencing

John H. Cochrane*

January 26, 2012

Empirical work in economics and finance frequently throws away a vast amount of the variation in the data. It is common to difference one, two, and even three times, in "difference-of-difference" estimates. For example, one might run a regression in which the data have been differenced over time, across states, and across industries or companies. In addition, it is common to include a multitude of dummies fixed effects, and controls, all of which remove variation from the data. The meaning of the regression coefficients changes apace, with the regression of left shoes on price and right shoes, and the regression of wages on education with industry controls as classic cautionary tales. (See ps at the end if these aren't obvious.)

There is a similar tension between robustness and efficiency. Maximum likelihood, GLS or other "efficient" procedures advocated by econometrics texts often suggest a lot of filtering of the data, or focusing on moments other than the straightforward correlations.

This note tells a little parable from money demand estimation to illustrate the dangers of over-differencing. The message is not that "differencing is bad." People usually difference or control in order to try to isolate exogenous variation when the right hand variables are endogenous. But this parable offers a graphic illustration of the possibility that the baby can be thrown out with the bathwater when much of the variation in the data is thrown out.

The standard theory of money demand holds that velocity $V = Y/M$ rises – money demand falls – when interest rates rise. When interest rates rise, the opportunity cost of holding cash rather than interest-bearing assets falls. Figure 1 shows that this prediction works remarkably well. (There is a question of what monetary aggregate to use and how to incorporate the vast expansion of highly liquid interest bearing assets. The St. Louis Fed's MZM definition tries to take account of this fact. The point here being econometric and not about the deep theory of money, I won't pursue the question.)

Figure 2 makes a scatterplot of the same data. We are soon going to run a regression of velocity on interest rates, and the eye can be deceiving in time-series plots, so this is a good check on a regression. (The eye tends to focus on the wiggles, where the theory works well

*University of Chicago Booth School of Business and NBER. john.cochrane@chicagobooth.edu, http://faculty.chicagobooth.edu/john.cochrane/

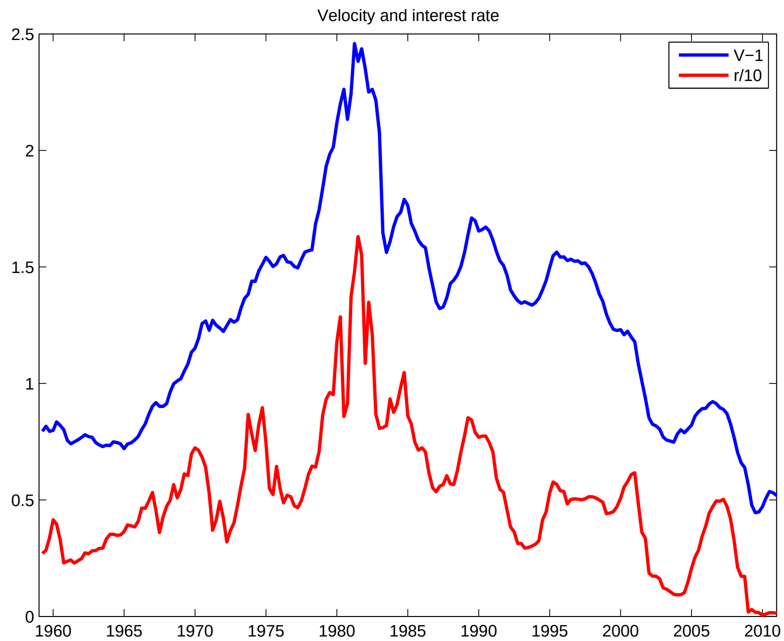

Figure 1: Velocity and interest rates. Velocity, the ratio of money to nominal GDP, is the St. Louis Fed MZMV series. Interest rate is the 3 month Treasury bill rate, TB3MS. I subtract 1 from velocity and divide interest rates by 10 so they fit on the same graph. Source: Fred database.

though the major source of variation is the big up and down. The eye also tends to shift time series to the right and left a bit in order to line up wiggles better, but regressions won't do that.)

				raw		nonpara.		parametric	
	b	R^2	ρ	se	t	se	t	se	t
levels	0.136	0.77	0.86	0.004	30.44	0.008	17.7	0.016	8.34
$x_t - \hat{\rho}x_{t-1}$	0.043	0.25	0.51	0.010	4.52				
$x_t - x_{t-1}$	0.018	0.09	0.31	0.005	3.37				

Table 1. Regressions of velocity on interest rates. 1959:1-2011:7 "Raw" standard errors are conventional OLS standard errors. "Nonparametric" use the Newey-West correction with 5 years of lags on each side. "Parametric" uses the correction factor $cov(\hat{\beta}) = (1+\rho)/(1-\rho)(X'X)^{-1}\sigma_\varepsilon^2$. The regression equations in each row are:

$$\begin{aligned} V_t &= a + br_t + \varepsilon_t \\ V_t - \hat{\rho}V_{t-1} &= a + b\left(r_t - \hat{\rho}r_{t-1}\right) + \left(\varepsilon_t - \hat{\rho}\varepsilon_{t-1}\right) \\ V_t - V_{t-1} &= a + b\left(r_t - r_{t-1}\right) + \left(\varepsilon_t - \varepsilon_{t-1}\right) \end{aligned}$$

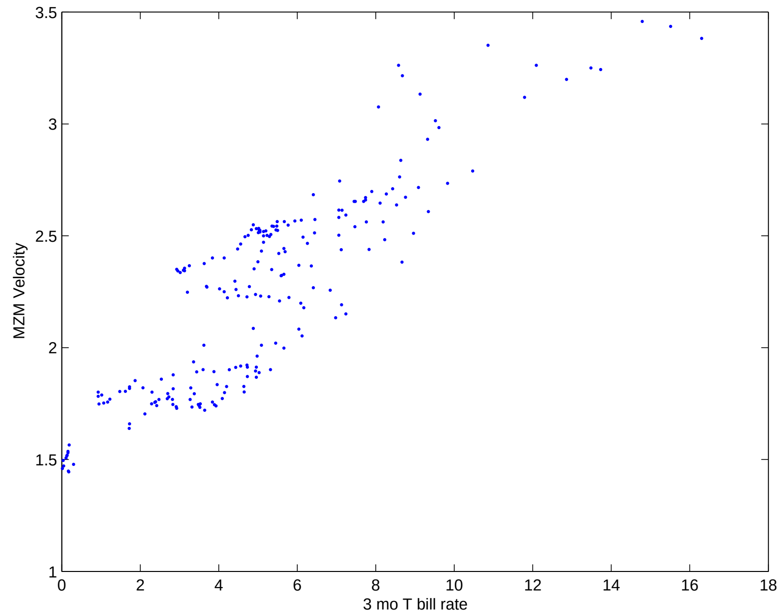

Figure 2: Velocty vs interest rate scatter plot.

Table 1, row 1, runs the regression of velocity on interest rates. We see a sensible coefficient, corresponding to standard views of the interest-elasticity of money demand. We have an impressive 0.77 R^2. It looks like a good regression. Except for the horrendous 0.86 autocorrelation of the residuals. As a result of that autocorrelation, the 30.44 t statistic is surely vastly inflated. The remaining columns give two better measures, which correct the OLS regression standard errors for the correlation of the residuals, giving much more sensible results. (Neither correction is aggressive enough, a lesson for another day.)

The standard econometric advice in this situation is to run GLS, or equivalently to quasi-first difference the data. OLS estimates are not biased, but GLS is efficient. Since $\rho = 0.86$, and is downward-biased at that, quasi-first differencing is essentially the same thing as first differencing.

The next two rows of Table 1 present the results of running the regression with differenced data. You see first of all a dramatic difference in the *economic* result. The slope coefficient declines by a factor of *ten* in the differenced specification. The R^2, once confirming the beautiful correlation in the plots, is down to 10%. The t statistics look more reasonable, which is good, but we know how to fix t statistics. And the residuals are still autocorrelated, suggesting even more differencing.

Figure 3 presents the differenced data, and Figure 4 presents the scatterplot. You can see that the lovely and persuasive correlation in the level picture is completely gone in the differences.

What happened? Why did the estimated coefficient change so much? When we transform

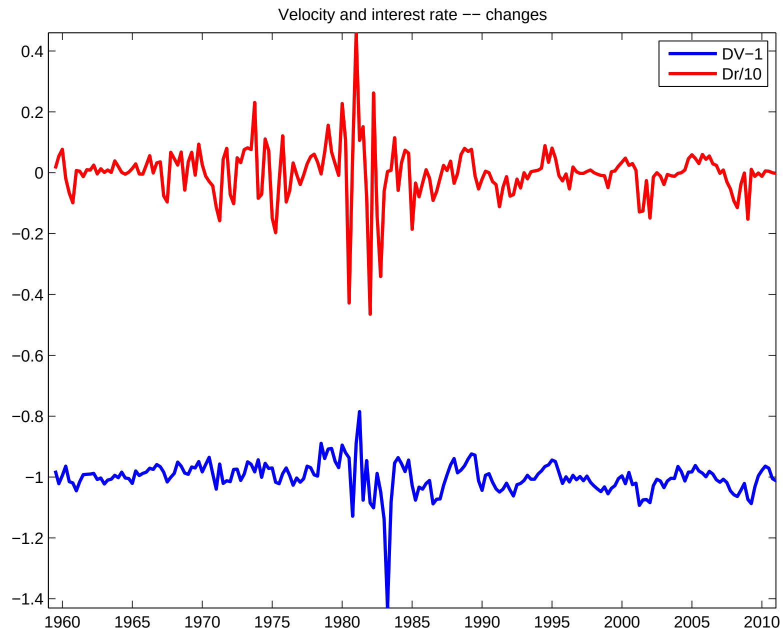

Figure 3: First difference of velocity, and interest rate. I subtract one from the first difference of velocity so the series can be visually distinguished.

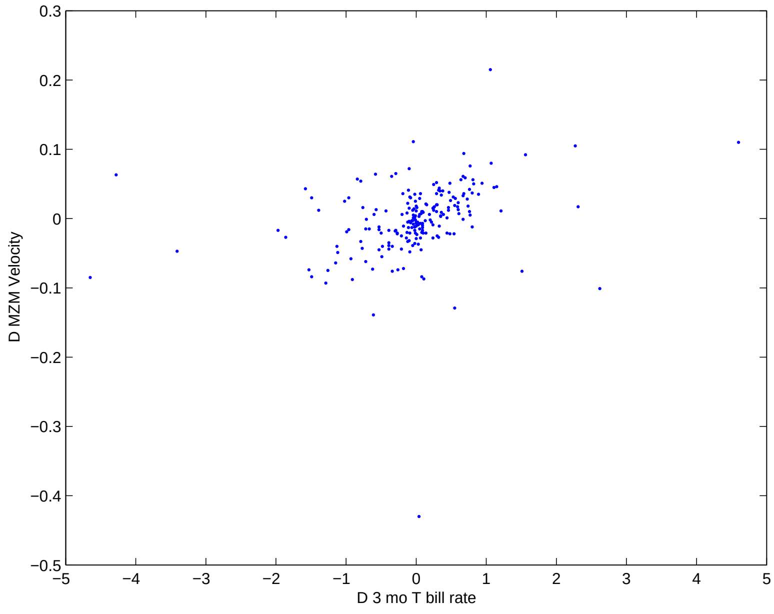

Figure 4: Scatter plot of changes in velocity and changes in interest rates

to differences, going from

$$V_t = a + br_t + \varepsilon_t \tag{1}$$

to

$$\Delta V_t = a + b\Delta r_t + \delta_t \tag{2}$$

we make a very important implicit assumption: we assume that the model is exactly right, and that the data are well measured. The "efficiency" of GLS relies heavily on those assumptions, to tell us that we can measure the b of equation (1) by measuring the b of equation (2) instead.

But no economic model is perfectly true, and no economic data are perfectly measured. There is surely some dynamic relationship between interest rates and money demand, at a minimum something like

$$V_t = a + b(L)r_t + \varepsilon_t. \tag{3}$$

In levels, we're basically measuring $b(1)$, while in differences, we're basically measuring $b(0)$.

The data are surely mis-measured. As a simple example a one-quarter lag in reporting either series would leave the level regression (1) pretty much intact, but destroy the differences regression (2). More realistically, suppose there is a small iid measurement error in the levels,

$$\begin{aligned} V_t^* &= V_t + v_t \\ r_t^* &= r_t + \rho_t \end{aligned} \tag{4}$$

The level regression (1) is relatively unaffected, but Δv_t and $\Delta \rho_t$ swamp the variation in the differences specification.

By differencing, then, we have tremendously emphasized the small specification and measurement errors in the data, relative to the signal that our theory is capable of describing.

A purist might say, ok, specify the dynamics in (3) and the measurement error process in (4), and let's go at it. But in this case as in many, economic theory really isn't up to explaining the high-frequency dynamics of money demand – if interest rates go up *this week*, how long does it take to adjust your cash management habits? Such a theory at a minimum needs to distinguish expected from unexpected interest rate changes, and address the fact that this is a very flat optimum in which near-rational decision rules cost pennies. Nor do we have any clue really what the physical and conceptual measurement errors are, let alone how to model them so maximum likelihood can go looking for efficiency. We're not *interested* in the dynamics or measurement error, we just want to see how the rough and ready velocity equation works.

For this reason, much research in macro and finance has given up on "efficient" estimation in favor of displaying easily digested and robust moments of the data (macro) or running OLS regressions and correcting standard errors in place of GLS regressions.

Again, differencing and dummies in panel data regressions are undertaken to find exogenous sources of variation, so "don't do it" is not good advice. But "ignore the limits of your model and mismeasurement of your data and throw away 99% of the variation in the data without concern" is not good advice either.

A good test is, "explain the source of variation in the data in words." In my case, you can see "higher interest rates are associated with lower money demand" is quite different from "a quarterly rise in interest rates is contemporaneously associated with a quarterly decline in money demand." Now, triple difference that, with 4 fixed effects, and 30 controls, and see if you can write the sentence.

Another good test is, "explain that your economic theory is well enough specified to describe the differenced and orthogonalized (to controls and fixed effects) data." That was clearly not the case in my example – $V = a + br$ is not a theory capable of describing quarterly *changes* in money demand, though it does a good job of the levels.

Finally, "describe the sources of specification and measurement error, and explain why differencing and orthogonalizing does not throw out more signal than noise." Again, the converse was the case in my example.

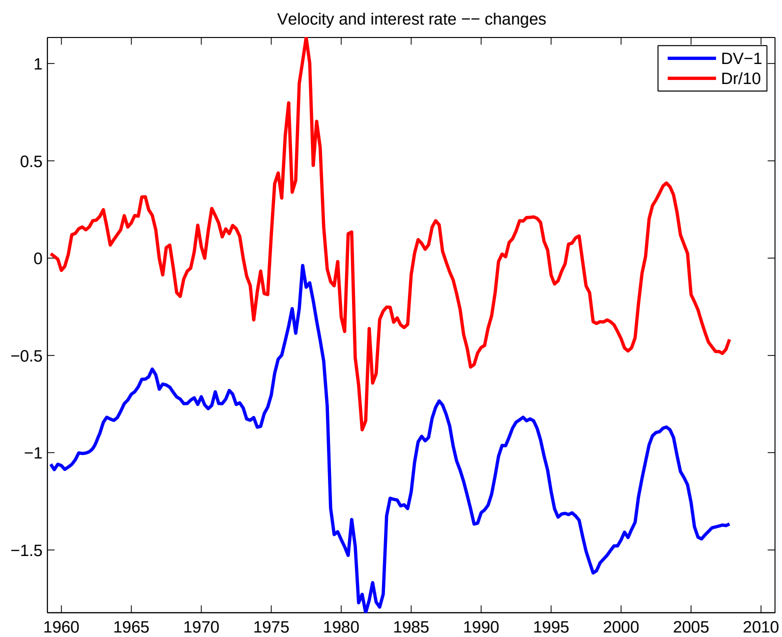

Figure 5: Four-year differences in velocity and interest rate

Not all differencing is bad. Sometimes the model does speak to the difference and not the level. Real business cycle models consciously are not models of fluctuation in "trend" growth, so they Hodrick-Prescott filter the data to focus on the correlations that they think their model is capable of addressing. Q theory works much better in differenced form than level form, though with annual not 10-minute differences. Sometimes it's the level that carries the misspecification.

In my example, the "level" regression suffers from a fault, that the source of variation is basically two data points, the rise and fall of inflation peaking in 1980. One might well want to know if the theory describes the cyclical wiggles in money demand, though seeing that it does not describe quarterly changes. Though filtering might be better, just to keep this note simple Figure 5 plots the four-year changes, as a way of looking at cyclical variation. You can see that this filter preserves enough signal relative to noise to answer that question.

Acknowledgement: This is not a new point either in the abstract or in the money-demand example. The example is motivated by Robert E. Lucas, 1988, "Money demand in the United States: A quantitative review," *Carnegie-Rochester Conference Series on Public Policy*, 29, 137-167. Bob figured out that the standard differenced specifications of money demand equations weren't working and that this was the right regression for measuring interest elasticity. He also rediscovered cointegration in estimating the income elasticity, and gave one of the best illustrations of the power of cointegrating-vector regressions.

PS: In case you forgot the classic fairy tales:

You're running a regression to estimate shoe demand, and you have data on left shoe sales. $Q = a - bP + \varepsilon$. A buddy notices the R^2 is pretty low, and suggests you add some controls or "explanatory variables." Aha, you have data on right shoe sales! $Q^L = a - bP + cQ^R + \varepsilon$. Now the R^2 is great and the t stats jump. Was this a good idea? No, because the coefficient now means, what is the effect of price holding constant right shoe sales! You have at best the demand curve for people missing right feet.

In a wage equation, $w = a + b \times educ. + \varepsilon$ generally produces R^2 in the 20% range. Ugh. Let's add more "explanatory variables." How about industry? That helps "explain" wages $w = a + b \times educ. + c \times industry + \varepsilon$ and your R^2 and t stats jump. But now the coefficient means, what if you get a PhD and stay in the fast food service industry! The point of education is to change industry, so you *don't* want to add this control. High R^2 is a bad sign!

Thanks to Tom Rothemberg, whose amazing econometrics class I still remember 30 years later.