



Alternative Approaches to Analyzing Economic Data

Milton Friedman; Anna J. Schwartz

The American Economic Review, Vol. 81, No. 1 (Mar., 1991), 39-49.

Stable URL:

<http://links.jstor.org/sici?sici=0002-8282%28199103%2981%3A1%3C39%3AAATAED%3E2.0.CO%3B2-T>

The American Economic Review is currently published by American Economic Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/aea.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Alternative Approaches to Analyzing Economic Data

By MILTON FRIEDMAN AND ANNA J. SCHWARTZ*

*This paper compares two approaches to the analysis of economic data: the exclusively econometric approach recommended by Hendry and Ericsson and the more eclectic approach that we employed in our books *Monetary History and Monetary Trends*. It concludes that the article by Hendry and Ericsson is mislabeled. Their article is not in any relevant sense an evaluation of our “empirical model of U.K. money demand.” Rather, it uses one equation from our book as a peg on which to hang an exposition of sophisticated econometric techniques. Insofar as their empirical findings do bear on ours, they simply confirm some of our principal results and contradict none. (JEL 036, 212)*

In response to the “on the one hand” and “on the other hand” economist, Frank H. Knight’s favorite remark was, “Yes, indeed, there are two sides to every issue: the right side and the wrong side.” David Hendry and Neil Ericsson (1991) (hereafter HE) take the same position—though in their case seriously not jocularly—about how to do empirical research in economics: HE’s way and the wrong way. We are more eclectic. We believe that there is no magic formula for wringing reasonable conjectures from refractory and inaccurate evidence. The HE approach is one way: start with a collection of numerical data bearing on the question under study, subject them to sophisticated econometric techniques, place great reliance on tests of significance, and end with a single hypothesis (equation), however complex, supposedly “encompassing”—to use one of HE’s favorite terms—all subhypotheses. Another way is to examine a wide variety of evidence, quantitative and nonquantitative, bearing on the question under study; test results from one body of evidence on other bodies, using econometric techniques as one tool in this process, and build up a collection of simple hypotheses that may or may not be readily viewed as components of a broader all-embracing

hypothesis; and, finally, test hypotheses on bodies of data other than those from which they were derived. Both ways—and no doubt still others as well—have their use. None, we believe, can be relied on exclusively.

I. A Comparison of Approaches

The wide difference between HE’s approach and ours is clear from HE’s section II, in which they “discuss the money-demand relationships they [we] estimated from the phase-average data for the United Kingdom” (HE, p. 9). From the literally hundreds of regressions in *Monetary Trends* (Friedman and Schwartz, 1982), they included four in their table 1 and concentrated their fire on only one of them. They assert that our regression for U.K. money demand is misspecified, “is not an adequate characterization of the data” (HE, p. 15) and does not support any of our “main claims about the United Kingdom” (HE, p. 32). They use that regression as a textbook example to illustrate a collection of sophisticated statistical techniques and to call the reader’s attention to the technical statistical literature dealing with those techniques—most of it published after our book went to press.¹ That may be a good way to intro-

*Friedman is Senior Research Fellow, Hoover Institution, Stanford University, and Professor Emeritus of Economics, University of Chicago; Schwartz is Research Associate, National Bureau of Economic Research.

¹Out of 108 total references in HE’s list, 70 were published after 1981, the year our book went to press; also, 20 out of 23 articles that were authored by Hendry alone or with collaborators.

duce students to the particular techniques covered and to the associated literature—which, we hasten to add, we are not competent to deploy, let alone judge—but it provides little if any evidence that is relevant to judging the validity of our analysis of money demand.

Hendry and Ericsson are entirely right that the one regression they subject to detailed scrutiny is not *by itself* adequate support for our “main claims about the United Kingdom” (see the Addendum to this article, in which Milton Friedman relates an experience that contributed to his skepticism about relying on a single regression produced by the HE approach). If that regression had been the only source of our “claims,” we would never have made them. As we emphasize in our book—partly in a passage that HE quote but also elsewhere—we do not have much confidence in “what has become the prevailing fashion in econometric work, the immediate computation of multiple regressions including all variables that can reasonably be regarded as relevant” (p. 215). The regression they analyze was selected from a table on page 282 (all unidentified page references herein are to Friedman and Schwartz [1982]) that contains six regressions, three for the United States and three for the United Kingdom, and was itself a prelude to presenting two equations “as a final summary of our results” for the United States and the United Kingdom combined (p. 284).

By HE’s standards, the prior 281 pages of our book are mostly worthless. Indeed, as we shall shortly demonstrate, while they may have skimmed those pages, they have apparently not thought it worthwhile to read them carefully. Those pages were not devoted, à la HE, to “representing the joint density of [a limited set of variables] in terms of an autoregressive-distributed lag model,” then proceeding to simplify “[t]he conditional model...to an ECM [error-correction model],” and to evaluating it “in light of the model design criteria” listed in their table 2 (HE, pp. 22–3). Instead, the first 204 of those 281 pages present our theoretical framework, our statistical framework, the basic data, and an overview of the

movements of money, income, and prices over the century our data cover. We end our broad overview by noting “three phenomena that require further interpretation: first, the common movement of nominal money and nominal income; second, the largely common movement in velocity in the United States and in the United Kingdom; third, the rather different relations in the two countries between movements in money, on the one hand, and real income and prices on the other” (pp. 184–5). Only then, in Chapter 6, “Velocity and the Demand for Money,” do we explore the first of the three phenomena, and, in subsequent chapters, the other two phenomena.

Throughout, our aim is to explain movements longer than a business cycle and to abstract from strictly intracyclical effects. That aim is both less and more ambitious than HE’s. It is less ambitious in that we limit attention to secular effects, while they seek a single econometric specification that simultaneously describes cyclical and secular movements. It is more ambitious in that we use as wide a range of evidence as we can, including qualitative examination of historical experience as well as numerical data for more than a century for both the United States and the United Kingdom. A major purpose of our book was to “give numerical content to some of the elements and hypotheses embedded in” “the broad theoretical framework that guided our earlier study of United States monetary history” “as well as some of the generalizations suggested by *A Monetary History*” (p. 13 *passim*). Accordingly, we frequently draw on evidence for the United States presented in the earlier book (Friedman and Schwartz, 1963). In particular, we use independent data for the United Kingdom to test generalizations initially suggested by U.S. experience.

The difference between HE’s approach and ours is exemplified by one of the test “criteria” by which they judge our equation unacceptable, namely that “[t]he values of $\hat{\sigma}$ [the standard error of estimate] for the corresponding subperiods [the first and second half, 1875–1928 and 1928–1975] are markedly different: 2.8 percent and 6.0 per-

cent, respectively” (HE, p. 13). On page 175 of our book, we noted that “For the United Kingdom, for the period before 1914, the extensive use of interpolation . . . biases sharply downward both the initial and the residual variation.” In consequence, if the error of estimate had been the same in the two periods HE compare, we would have regarded that fact as evidence *against*, not for, our equation. That would have meant that we had erroneously attributed pure measurement error to economic variables. Similarly, the finding by HE that the equations they prefer (their equations 10 and 11) have the same residual variability before and after 1914 (not 1928) is, in our view, evidence *against*, not for, their equations.

They say, we “do not formally *test* for constancy” (HE, p. 13) by which they mean we do not use the particular statistical tests that they advocate; but we certainly do test for constancy by comparing regressions for different periods, trying to isolate features of the statistical data that may bias comparisons, drawing on historical data, and in a variety of other ways. We regard such tests as far more reliable and, to use a statistical term, robust than the formal statistical tests used by HE.

Another example is their assertion that “price homogeneity” incorporated in our model is rejected by their statistical test and their accompanying comment that, when we “tested for trends and price homogeneity in more restrictive models, they [we] did not obtain rejections” (HE, p. 15). From our viewpoint, the tests for price *and population* homogeneity on pages 253–9 were less, not more, restrictive than HE’s test.² Our tests were for both the United States and the United Kingdom, used both levels and rates of change, and allowed for the “regression effect,” of which more later, as well as for a variety of possible explanatory variables. We concluded that the results were “reasonably consistent with the” assumption that “any change in prices or population implied an equal percentage change in the quantity of

money demanded (mathematically, the demand for money in nominal terms is homogeneous of the first degree in population and prices)” (pp. 258, 253). We hasten to add that our conclusion was for a long-run demand function, not for a function seeking to cover intracyclical effects.

Still another example is HE’s rejection of our use of phase averages and their decision to analyze our “original annual observations” (HE, p. 18). That is a correct decision for their objective: a single equation that “encompasses” both secular and cyclical effects. However, that was not our objective. We devoted pages 79–97 to testing whether the phase averages and rates of change contain a residual cyclical element and whether they are infected by a spurious serial correlation. We concluded “that our phase bases do eliminate the bulk of the systematic cyclical fluctuation” and “that the ‘noise’ introduced by the serial correlations is sufficiently small relative to the systematic variation we are trying to describe that it can for the most part be neglected” (pp. 81, 97). True enough, we discovered, as HE do, that annual observations give essentially the same results as phase-average data for the regression they analyze, and its counterpart for the United States (see Friedman, 1988 footnote 16). However, we doubt that we would have settled on those specific final equations if we had relied solely on annual data in our prior exploratory investigation. Throughout that investigation, we did not have to allow for cyclical effects, or explore a “(dynamic) error-correction representation of the relevant variables” (HE, p. 20), as we would have had to do if we had followed HE’s prescription. Hence, we remain unpersuaded that our use of phase-average data was a mistake.

II. Exogeneity

Hendry and Ericsson put much emphasis on “concepts of exogeneity,” listing four distinct concepts, and stating: “In no case is it legitimate to ‘make variables exogenous’ simply by not modeling them” (HE, p. 21). We do not sympathize with such commandments. In our view, exogeneity is not an

²So far as we can see, HE nowhere even try to test for population homogeneity.

invariant statistical characteristic of variables. Everything depends on the purpose. In economic analysis, it may be appropriate to regard a variable as exogenous for some purposes and as endogenous for others. A simple example is the quantity of money. For the United States after World War I, we believe it is appropriate to regard the money stock as exogenous (i.e., determined by the monetary authorities) in an economic analysis of long-run money demand. We do not believe it would be equally appropriate to do so for week-to-week or month-to-month movements for which, in HE's words, "the money stock appears to be endogenously determined by the decisions of the private sector" (HE, p. 27). For the period before World War I, as we repeatedly state in both *Monetary History* and *Monetary Trends*, it is not appropriate to regard the money stock as exogenous even for money-demand analysis, particularly for phase averages. For that period, the money stock is best regarded as endogenous, because the United States and much of the rest of the world was on a gold standard. Even for the post-World War I period, it would not be appropriate to regard the money stock as exogenous in a broader study of the Federal Reserve as a political institution, created by Congress, subject to its ultimate control, with members of its Board appointed by the President. For such studies, of which there are a good number, the money stock is best regarded as endogenous.

After their own examination of "exogeneity" and "super exogeneity" for one of their regressions, HE assert that "This evidence is inconsistent with the hypothesis that, over the period 1878–1970, exogenous money determined prices in the United Kingdom via a stable demand function, *precisely because we have established a constant money-demand model conditional on prices*" (HE, p. 30). We interpret their italics as implying that we asserted the hypothesis that they regard as contradicted. We clearly did nothing of the kind. In particular, a large part of our next chapter (Ch. 7) explores the direction of influence of the variables HE consider as well as the direction of influence between the United States and the United

Kingdom. We distinguish different monetary regimes and examine separately the gold-standard period and the variable-exchange-rate period. At the outset of Chapter 7, we note that "one possible explanation for the similar behavior of [velocity] is that the two countries were part of a single economic entity" (p. 305). We go on to note that before World War I, when both countries were on a gold or sterling standard, "as we demonstrated in *A Monetary History*, ... there was a good deal of leeway for domestic monetary policy over short periods, but over periods of more than a few years, the quantity of money in each country was determined by the requirement that the price levels of the two countries move roughly in step in order to preserve equilibrium in the balance of payments" (p. 306). Clearly, we are not guilty of the sin that we suspect, perhaps wrongly, HE accuse us of.

III. The Regression Effect

On the purely statistical level, we tried throughout to avoid making arbitrary assumptions about exogeneity. For example, in the first table in Chapter 6 (table 6.1, p. 212), we label one set of equations "Level of Money: Income Assumed Exogenous" and another "Level of Income: Money Assumed Exogenous" and do the same for rate-of-change calculations. Similar caution in later tables derives in large part from our trying systematically to allow for what we referred to earlier as "the regression effect."

The regression effect refers to the bias introduced by stochastic disturbances ("error") affecting a variable treated as exogenous. Indeed, regression analysis got its name from this bias.³ Put formally, consider a relation between two variables (x and y) that is strictly linear. Given errors of mea-

³In studying the relation between the heights of fathers and sons, Sir Francis Galton discovered that the average height of sons of tall (short) fathers "regressed" to the mean (i.e., differed from the average height of all sons by less than their fathers' height differed from the average height of all fathers) and, simultaneously, the average height of the fathers of tall (short) sons also "regressed to the mean."

surement, a sample of observed values of the two variables will yield a bivariate scatter, rather than a strictly linear relation. The calculated regression of y on x will be flatter than the "true" relation, and that of x on y will be steeper. Harold Hotelling pointed out in 1933 how potent a source of economic fallacies the regression effect can be if "errors of measurement" are interpreted to include all stochastic disturbances affecting the variables under study (Hotelling, 1933, 1934; Horace Secrist, 1934; Friedman and Simon Kuznets, 1945 pp. 325–38; Friedman, 1957 Ch. 3).

A major difference between our statistical approach and HE's is that we allow systematically for the regression effect. HE obliquely recognize this difference in their data appendix, where they note near the end that we "emphasize an 'errors-in-variables' paradigm" (HE, p. 34). Nowhere else do they mention or refer to the regression effect or seek to allow for it. In our opinion, it is often a more important source of error—because it introduces systematic bias—than is the residual error described by the standard error of estimate to which HE give exclusive attention.

In our analysis, as in many economic analyses, the so-called "independent" or "exogenous" or "predetermined" variables are infected by error in a double sense. In the first place, they are not always precise empirical counterparts to the economic variables that theory suggests including. To illustrate, we regard the yield on physical assets as one of the interest rates that can be expected to affect the real quantity of money demanded. However, despite considerable exploration, we could not find a direct empirical counterpart. We finally settled on using the rate of change of nominal income as a "proxy for the nominal return on physical assets" (pp. 274–80). As a second example, we consider as another relevant variable, the own yield on money, to which HE refer in the final paragraph of their section IV, citing a 1987 article, but do not empirically explore. Here again, we were unable to find a reliable reasonably direct measure, though further statistical exploration of basic data may enable such a mea-

sure to be constructed. Instead, we use an indirect measure suggested by Benjamin Klein (1970) and described by HE in their footnote 26. (We examine the whole issue on pp. 259–73.) In both cases, even if the proxy were measured precisely, there would remain measurement error equal to the difference between the relevant economic variable and the proxy used as its empirical counterpart.

In the second place, the proxy variable is never measured precisely, so there is a measurement error in the ordinary sense that adds to the total measurement error. When these measured variables are used in a regression as independent variables, their coefficients are biased estimates of the underlying theoretical coefficients. One way to estimate the possible size of the bias is to reverse the direction of regression: in the bivariate case, to estimate both a y -on- x regression and an x -on- y regression. The resulting estimates of the theoretical coefficients provide an upper and a lower limit. A similar procedure can be used to get upper and lower limits in multiple regressions. We discussed this issue at some length and dealt with it by consistently calculating such limits for essentially all of the parameters we estimated.⁴ In principle, the limits should be made still wider to allow for the usual standard error of estimate of the parameters arising from the stochastic element in the dependent variable. We did not do so; however, we consistently presented t ratios, from which what is generally called the "sampling error" can be readily calculated, so that any interested reader can combine the two.

An example shows the importance of the regression effect compared with the conventional "sampling error" that is HE's sole concern. Consider the coefficient of the interest rate that HE refer to in their section

⁴See especially Chapter 5, footnotes 28 and 29, pp. 173–4; Chapter 6, footnote 7, p. 211, footnote b to table 6.1, p. 213, footnote 18, pp. 224–5; and the text to which these footnotes are attached, as well as comments scattered throughout the text in discussing particular sets of results.

II and plot in their figure 5. Our upper and lower point estimates allowing for the regression effect differ by 28.6 (table 6.15, p. 285). One estimate of the standard error of the coefficient is 3.26.⁵ Even four times the standard error (plus or minus twice the standard error) is 13.04, or less than half the regression effect. Combining the sampling error and the regression effect gives a range for the coefficient from -4.6 to -46.3 , the basis for our statement that this coefficient was “much less precisely estimated” than the income elasticity, for which the regression effect was 0.08, the standard error of estimate of the parameter 0.005, so the combined range was 0.87–0.97.

This example is, we believe, fairly representative with respect to the relative importance of the regression effect and the regularly computed standard error of estimate of parameters. It suggests why we attribute so much importance to the regression effect, or alternatively, why we regard the “errors-in-variable model” as appropriate for economic data that are subject to large differences between the theoretical variables and their statistical counterparts and to large measurement errors.⁶

IV. The Proof of the Pudding

Hendry and Ericsson’s final products which, by comparison with our results, they describe as a “better-fitting, constant, dynamic, error-correction (cointegration) model” (HE, p. 8) and which satisfy all of

their statistical “criteria for evaluating and designing models” (HE, p. 19) are their regressions 10 and 11, the first, with 10 parameters, omitting the five final years, 1971–1975; the second, with 12 parameters, covering the whole period, 1878–1975. These are in turn simplifications of the “general autoregressive-distributed lag” regression in their table 3 containing 37 parameters and omitting the final five years. How do their results compare with ours? In terms of the long-run effects we were interested in, they simply confirm a few of our results.

A. *Income Elasticity of Money Demand*

Our point estimate of the elasticity of per capita real demand is 0.88; allowing for sampling error by adding plus or minus twice the standard error of estimate of the relevant parameter gives a range of 0.87–0.89; allowing also for the regression effect gives, as noted earlier, a range of 0.87–0.97. The regression yielding these estimates has six parameters.

Neither HE’s regression 10 nor their regression 11 contains an explicit income term. The reason is presumably their prior conclusion, embodied in regression 9 and apparently regarded by them as confirmed by the 37-parameter regression in their table 3, that the income elasticity of aggregate money demand is unity. Nowhere do they allow for population (unless implicitly in the constant terms of regressions 10 and 11). Since we decided that the elasticity of money demand with respect to population could be taken as unity, the aggregate elasticity can be expected to be closer to unity than the per capita elasticity. The regression in HE’s table 3 does include income separately and so permits an elasticity other than one. If we convert it into a long-run regression by adding the coefficients of the separate lag terms, as is done in the final column of the table, the result is an eight-parameter regression and a point estimate of the income elasticity of 1.1. Clearly there is no contradiction between their estimate and ours if allowance is made for both sampling error and the regression effect.

⁵“One estimate” because it is from one of the two regressions from which we calculate the limits (e.g., y on x). The other regression (e.g., x on y) would give a somewhat different estimate.

⁶In re measurement error, in a later chapter (Ch. 8), we use a highly indirect approach to estimate the pure measurement error in our data. For the phase-average level of nominal income, we estimate that the standard deviation of pure measurement error is 2.5 percent for the United States, 4.5 percent for the United Kingdom. For the rate of change of nominal income computed from the phase averages, we estimate the corresponding standard deviations to be 0.74 of a percentage point for the United States and 0.58 of a percentage point for the United Kingdom (pp. 350–1). These are appreciable, yet not unreasonable.

B. *Interest Semielasticity of Money Demand*

Our point estimate of the semielasticity of money demand with respect to the differential yield on money (the short-term interest rate minus our proxy for the own-rate on money) is -11.16 . This implies that a *one percentage point* increase in the differential interest rate reduces the real quantity of money demanded per capita by 11.16 percent. As noted earlier, we regard this estimate as not very well determined and the range, including both sampling and regression effects, is extremely wide, from -4.6 to -46.3 , though, as theory would lead one to expect, negative throughout.

HE produce no strictly comparable elasticity. They use the short-term interest rate but do not allow for the own rate.⁷ In commenting on their equation 9, they state that it “implies that a one-percentage-point increase in the short-term interest rate... reduces [velocity]... by 7 percent in the long run,” (p. 25) or that their estimate of the semielasticity of the short-term interest rate is -7 , well within our estimated range.⁸

C. *Postwar Readjustment and Demand Shift*

Our regression includes two dummy variables to allow for postwar readjustment and a demand shift. HE remark that “many investigators would regard the need for the data-based shift dummy \bar{S} spanning one-

third of the sample as *prima facie* evidence against the model’s constancy” (HE, p. 13). However, the regression in their table 3 has three dummy variables, one of which is the precise counterpart of our demand shift dummy \bar{S} ; the other two, like our postwar adjustment dummy, allow for wartime effects. Their regression 10 combines the two wartime dummies but does not explicitly use a demand-shift dummy. Their regression 11 introduces an additional dummy for the 1971–1975 period.

The coefficient of our demand-shift dummy is 0.21, implying that the demand for money shifted upward by 21 percent during the period in question. Adding and subtracting two standard errors of estimate gives a range of 0.15–0.27. HE’s table 3 gives a point estimate, in units comparable to our estimate, of 0.06 but with a very large standard error, so that adding and subtracting two standard errors gives a range from -0.29 to $+0.41$. Commenting on their “simplified” regression 10, HE note that it “exhibits multiple equilibria, with two corresponding to the long-run solution (9) and a third being that solution shifted by 20 percent. In that sense, the results are consistent with the use of an adjustment factor of about that order [i.e., 21 percent] by Friedman and Schwartz” (HE, p. 26). We have not been able to devise any ready way to compare our estimate of the postwar readjustment effect with the effect of their two wartime dummies.

D. *Differences in Scope*

So far, the HE results simply confirm ours. In addition, we 1) estimated the effect of the proxy yield on physical assets, a variable they do not consider; 2) allowed for the effect of the own-yield on money, which they postpone for future work; 3) compared demand functions in the United States and the United Kingdom, concluding that the only important difference was a higher income elasticity for the United States than for the United Kingdom; and 4) estimated two regressions for the two countries combined, including one additional dummy to allow for the difference in income elasticity.

⁷However, in their footnote 26, HE assert that they experimented with using our proxy and that “The resulting estimates are very similar to those found using” only the short-term rate.

⁸In one of our preliminary regressions that HE do not refer to, the point estimate of the semielasticity for the short-term rate alone is -3.0 ; the range, allowing for both sampling error and the regression effect, is from -1.4 to -11.1 , again including their -7 (see tables 6.12 and 6.N.3, pp. 272–3 and 277).

The long-run regression derived from HE’s table 3 contains both a short-term and a long-term interest rate. However, as judged from the final column, the coefficient of neither comes close to being significantly different statistically from 0, and any range constructed from those point estimates is so wide as to be meaningless.

ties. HE, as part of their more ambitious agenda, estimate the short-run error-correction process, something that we explicitly excluded from our study.

E. *Encompassing*

Hendry and Ericsson put a great deal of emphasis on whether one "model" encompasses another. They assert that "a necessary condition for encompassing is *variance dominance*, where one equation variance-dominates another if the former has a smaller error variance" (HE, p. 22). A footnote attached to this sentence states: "Formally, variance dominance refers to the underlying (and unknown) error variances. Without loss of clarity, we often will say a model variance-dominates another if the *estimated residual* variance of the former is smaller than of the latter." In other words, they judge the validity of their hypotheses by the data from which they derive them!

In this spirit, they boast that "the error variance of (10) is less than one-tenth of that in" our regression (HE, p. 27); but this is to compare apples and oranges. Their regression 10 has a first difference or a rate of change as a dependent variable. The regression they compare it with has the logarithm of the level of money as the dependent variable. They present only two regressions in log levels: their regression 9 and the regression in their table 3. The standard error of estimate for regression 9 is nearly twice the standard error of estimate of our regression; for the regression in their table 3, it is nearly three times higher.⁹

We also estimated a parallel regression in rates of change, comparable to HE's regressions 10 and 11. It has six parameters and a standard error of estimate of 1.34, compared with 1.424 for their 10-parameter regression 10, which covers a shorter period than ours, and 1.478 for their 12-parameter

regression 11, which covers the same period as ours. In their terms, our regression variance-dominates theirs. They obliquely recognize the problem and try to justify their procedure: "Although (10) has a rate of change as the dependent variable, it is an equation in log-levels because of the error-correction term... , so direct comparison [with our log-level equation] is valid" (HE, p. 27). However, their "error-correction terms" are useful primarily for short-run analysis of levels, not for the long-run effects that were our sole concern.

HE's regressions 10 and 11 are misleading for a very different reason as well. The same variable, Δp_t , is included on both sides of the equation: as part of the dependent variable and also as an independent variable. Adding Δp_t to both sides of the regression does not change its statistical characteristics ($\hat{\sigma}$, the standard errors of the parameters, and the coefficients of all terms other than Δp_t , are unchanged). What it does make clear is that the regression "explains" a change in the nominal quantity of money, not the real quantity of money (is it, perhaps, best interpreted as an estimate of a short-run supply curve of nominal money?).

We have been unable to duplicate regression 10 precisely, but we have come reasonably close. Omitting Δp_t from the right-hand side nearly doubles the standard error of estimate (2.83 percent versus 1.50 percent).¹⁰ A parallel equation with Δm_t as the independent variable, and omitting Δp_t , gives a standard error of estimate one-third larger than that from our estimate of regression 10 (2.07 percent versus 1.50 percent). That is a reasonably satisfactory empirical equation for the change in the nominal quantity of money. However, we are hard put to construct any satisfactory theoretical interpretation of the regression.

HE refer favorably to papers by Andrew Longbottom and Sean Holly (1985a) and by Alvaro Escribano (1985), which they regard

⁹The standard error of estimate for our regression is 5.54 percent; for HE's regression 9, it is 10.86 percent; for that in their table 3, it is 15.5 percent (it is given as 1.55, but that has to be multiplied by 10 to be comparable to the other two, since the long-run dependent variable is $0.1m$).

¹⁰HE report the standard error of estimate as 1.424 percent. Our attempted duplicate has a standard error of estimate of 1.50 percent.

as producing “significant improvements on our [i.e., HE’s] 1983 model” (HE, p. 24) but still leaving room for further improvement, as in HE’s revised models. In a slightly later paper by Longbottom and Holly (1985b), which HE do not refer to but which, like the earlier paper, uses HE’s proposed procedures, the final summary is: “In this paper we have re-examined Friedman and Schwartz’s work on UK monetary trends in the light of the methodological criticisms of Hendry and Ericsson. In contrast to Hendry and Ericsson we are able to find empirical support for the claims that Friedman and Schwartz make about the form and long run stability of the demand for money function in the UK since 1878” (Longbottom and Holly, 1985b p. 19).

We hasten to add that we regard the Longbottom and Holly papers no less an example of formal econometric analysis carried to extremes than HE’s papers. Their confirmation of our results does not increase our confidence in our results any more than HE’s assertion that we have failed “to present statistical evidence pertinent to their [our] main claims about the United Kingdom” (p. 32) weakens our confidence in them. Our confidence in our results derives, as we have stressed repeatedly, from a much broader base and would not be justified if their analysis was all we had to rely on.

F. *The Real Proof of the Pudding*

Before ending this section, in which we have expressed so much skepticism about HE’s approach, we should indicate what evidence would persuade us that we are wrong and they are right. The answer is straightforward. A persuasive test of their results must be based on data not used in the derivation of their equations. That might mean using their equations to predict the same kind of phenomena for other countries, or for a future or earlier period for the United Kingdom, or deriving testable implications from their equations for other variables, such as exchange rates, term structure of interest rates, or still other phenomena we are not imaginative enough to list. Similarly, that is the only kind of evi-

dence that we would regard as persuasive with respect to the validity of our own results.¹¹

V. Conclusion

Hendry and Ericsson describe their paper as an evaluation of “an empirical model of U.K. money demand developed by Friedman and Schwartz in *Monetary Trends*” (HE, p. 8). Viewed from that point of view, seldom can a mountain have labored so hard and produced so small a mouse. After years of experiments, HE’s econometric techniques produced a series of models that confirm some of our principal results, contradict none, and are less successful than our equations in terms of their own criterion of variance-dominance.

But their paper is mislabeled. It is not in any relevant sense an evaluation of our “empirical model of U.K. money demand.” They use one out of our hundreds of regressions as a peg on which to hang an exposition of a set of sophisticated econometric techniques designed for a purpose and embodying a methodological approach very different from ours. Their regressions are designed to explain the short-term adjustment process as well as the long-term relation. We had no such ambitious aim.

We are incompetent to judge the adequacy of their techniques for estimating the short-term adjustment process. However, we do not regard their statistical tests as demonstrating the validity of their statistical estimates. Their estimates are the end result of trying a large number of alternative hypotheses on a single body of data. As a result, it is impossible to specify how many

¹¹Our initial draft of *Monetary Trends* was based on U.S. data only. We expanded it to include the United Kingdom precisely in order to test our initial generalizations with data other than those used in deriving them. This is of course a never-ending process; the generalizations in the published book in their turn should be tested against data not used in deriving them. One minor such test extending the period originally covered is that in Friedman (1988 footnote 16). A regression for the United States for 1886–1985 using annual data gives essentially the same results as our phase-average equation for 1873–1975.

“degrees of freedom” have been used up in the process of reaching the final equations presented, or, put differently, to estimate the probabilities that their results could have arisen from chance. For that, one needs, in their words, “the underlying (and unknown) error variances,” not “the *estimated residual variance*” on which they rely (HE, footnote 14). As already indicated, the real proof of their pudding is whether it produces a satisfactory explanation of data not used in baking it—data for subsequent or earlier years, for other countries, or for other variables. One example of such a test, in a physical-science context, is given in the Appendix. That example dramatically illustrates how misleading a multiple regression can be for predictive purposes, even though it satisfies all the standard tests.

APPENDIX: A CAUTIONARY TALE ABOUT
MULTIPLE REGRESSIONS

*(This addendum was written by
Milton Friedman)*

My skepticism about relying on a single multiple regression that results from the HE approach traces back to an experience I had in 1944 or 1945 when I was engaged in war research as a member of the staff of the Statistical Research Group of Columbia University.

One of my assignments was to serve as a statistical consultant to a number of projects seeking to develop an improved alloy for use in airplane turbo-superchargers and as a lining for jet engines. The goal was to develop alloys that could withstand the highest possible temperature, since the efficiency of a turbine (or its equivalent) rises very rapidly with the temperature at which it can safely operate. I served as something of a clearing agency for the results of the various experiments in progress, as an adviser on statistical design of experiments, and as an analyst of the results, producing a fairly regular newsletter on these matters for the experimenters.

The procedure in testing an experimental alloy was to hang a specified weight on a standard turbine blade made from the alloy,

put it in a furnace capable of generating a very high temperature, and measure the time it took for the blade to break. At one point, I combined the test data from all the separate experiments and engaged in precisely the kind of analysis that HE recommend. I ended up with a single proposed regression that expressed time to fracture as a function of stress, temperature, and variables describing the composition of the alloy. I assured myself that the equation was consistent with metallurgical theory.

The major problem then, trivial now, was to compute the parameters of the equation and the associated test statistics. That was the age of the desk electric—not electronic—calculators and the Dolittle method of computing regressions. The labor involved in that method increases exponentially with the number of independent variables. For the number I wanted to use, we estimated that it would take three months for one of our highly skilled operators to calculate the equation. Fortunately, we discovered that there was one large-scale computer in the country that could perform our calculations: the experimental Mark I (or something like that) at Harvard, itself not electronic but built from a large number of IBM card-sorting machines housed in an enormous air-conditioned gymnasium. We were granted time on the machine to perform our calculations. Today’s statisticians will be interested to know that, not counting data insertion, it took 40 hours to calculate a regression that I can now calculate on my desktop computer in less than 30 seconds—my favorite story to illustrate what has happened to our computer power.

I was delighted with the calculated regression. It had a high multiple correlation, low standard error of estimate, and high t values for all of the coefficients, and it satisfied every other test statistic that I knew of more than 40 years ago. I immediately set to work to create some new and better alloys. In constructing such alloys, I had to go outside the joint range of my sample set of independent variables, but I was careful to stay as close as I could and to be within the limits used in prior experiments for each variable separately. The technical details are

irrelevant for the present purpose, and I could no longer reproduce them in any event.

The bottom line is that I ended up constructing two new alloys (which with hope combined with caution, I named F-1 and F-2). According to the calculated regression, each would take several hundred hours to rupture at the very high temperature I proposed to test them at, a sizable multiple of the best recorded time for any previous alloy. This was physics, not economics, so I did not have to wait years to see whether the predictions from my equation were correct. I phoned an MIT lab that was working on alloys of a similar type and asked them to cook up and test my two alloys. I was sufficiently skeptical—or perhaps just cautious—so that I was careful not to tell them what to expect. A few days later they phoned the results: my two alloys had ruptured in something like 1–4 hours, a much poorer outcome than for many prior alloys. F-1 and F-2 were never heard of again.

Ever since, I have been extremely skeptical of relying on projections from a multiple regression, however well it performs on the body of data from which it is derived; and the more complex the regression, the more skeptical I am. In the course of decades, that skepticism has been justified time and again. In my view, regression analysis is a good tool for deriving hypotheses. But any hypothesis must be tested with data or non-quantitative evidence other than that used in deriving the regression or available when the regression was derived. Low standard errors of estimate, high t values, and the like are often tributes to the ingenuity and tenacity of the statistician rather than reliable evidence of the ability of the regression to predict data not used in constructing it.

REFERENCES

- Escribano, Alvaro, "Non-linear Error-Correction: The Case of Money Demand in the U.K. (1878–1970)," mimeo, University of California, San Diego, 1985.
- Friedman, Milton, *A Theory of the Consumption Function*, Princeton: Princeton University Press, 1957.
- _____, "Money and the Stock Market," *Journal of Political Economy*, April 1988, 96, 221–45.
- _____, and Kuznets, Simon, *Income from Independent Professional Practice*, New York: National Bureau of Economic Research, 1945.
- _____, and Schwartz, Anna, J., *A Monetary History of the United States, 1867–1960*, Princeton: Princeton University Press (for the National Bureau of Economic Research), 1963.
- _____, and _____, *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975*, Chicago: University of Chicago Press (for the National Bureau of Economic Research), 1982.
- Hendry, David F. and Ericsson, Neil R., "An Econometric Analysis of U.K. Money Demand in *Monetary Trends in the United States and the United Kingdom* by Milton Friedman and Anna J. Schwartz," *American Economic Review*, March 1991, 81, 8–38.
- Hotelling, Harold, "Review of *The Triumph of Mediocrity in Business* by Horace Secrist," *Journal of the American Statistical Association*, December 1933, 28, 463–5.
- _____, Letter to the Editor, *Journal of the American Statistical Association*, June 1934, 29, 198–9.
- Klein, Benjamin, "The Payment of Interest on Commercial Bank Deposits and the Price of Money: A Study of the Demand for Money," unpublished Ph.D. Dissertation, University of Chicago, 1970.
- Longbottom, Andrew and Holly, Sean, (1985a) "Econometric Methodology and Monetarism: Professor Friedman and Professor Hendry on the Demand for Money," London Business School Discussion Paper No. 131, February 1985.
- _____, and _____, (1985b) "Monetary Trends in the UK: A Reappraisal of the Demand for Money," London Business School Discussion Paper No. 147, April 1985.
- Secrist, Horace, Letters to the Editor, *Journal of the American Statistical Association*, June 1934, 29, 196–8, 200.