The Remarkable Long-Run Conditional Predictability of US Real M1

Clinton A. Greene*
University of Missouri-St. Louis

January 4, 2011

Abstract

A static cointegration model of US M1 forecasts beyond two decades ahead, conditional on known non-M1 variables. Data is adjusted for misreporting induced by regulatory accounting "sweeps". K'th-quarter-ahead forecasts from this model have smaller RMSE than those from a univariate (differenced) ADL model for k >5. The model employs money and income scaled per household. There are three theoretical reasons for doing so. Scaling is necessary (a priori) to avoid inducing instability, miss-timing and trivial "self-cointegration". The approach can be rearranged to model the price-level. Potential for long-run inflation forecasting (conditioned on other variables) is explored.

JEL: E41, E47, E31, E37

Keywords: money demand, M1, sweeps, scaling, forecasts, inflation

---

* Correspondence: Clinton A. Greene, Department of Economics, University of Missouri-St. Louis, St. Louis, MO, 63121-4499. Tel: (314) 516-5565. Fax: (314) 516-5352. Email: clinton_greene@umsl.edu. http://www.umsl.edu/~greenec/

0. Introduction

This paper shows a simple long-run model of US real M1 forecasts remarkably well, even twenty-five years ahead. The initial forecasting exercise is severe, estimating a static levels model over quarterly data 1959-81.3 and then using these estimates (without updating) to forecast through 2008, taking income and interest rates as known during the forecast period.[1] At 1981.3 there is a sharp reversal in trend velocity, followed a year later by a large and sustained shift in trend real money. Terminating the estimation sample in 1981.3 ensures the model must forecast the large shift in trend, however one prefers to measure the shift. The long-run cointegrating model also does remarkably well in shorter run forecasting, beating differenced univariate (ADL) models at time horizons of six quarters or more.[2]

The model is distinguished from other approaches in scaling money and real GDP on a per household basis. Section 1 discusses the theoretical reasons for scaling in this manner, and also presents a bit of historical background. Section 2 discusses the M1 data and the adjustments needed after 1993 for a measure which includes most freely checkable explicitly insured deposits. Section 3 shows the results of the long-run forecasting exercise, comparing the model scaled on a per household basis to the corresponding non-scaled model.

The remaining sections focus on scaled (per household) models. Section 4 examines integration and cointegration tests, and forecast performance based upon estimation methods other than Engle-Granger OLS. Section 5 presents short-run forecasting performance, comparing the static per household model to differenced univariate models in recursive

---

[1] These are "conditional pseudo forecasts", since only the dependent variable is treated as unknown.

[2] Lag truncation for the time-series model is re-evaluated at each step, hence I refer to these forecasts as coming from a series of time series models rather than from a single model.

forecasting from one through twenty quarters ahead. Section 6 introduces obstacles to and prospects for inverting the model for price-level and inflation modeling.

This paper adopts the view that the most fundamental task is to prove the usefulness of the long run equilibrium relationship. There is a long history of failures in theory-based forecasting and of apparent breakdown of money models, from Goldfeld (1976) to Hess Jones and Porter (1998). Given this history, results dependent upon tests with restrictive assumptions can reasonably be taken as weak evidence. If one finds this paper's results for the static cointegrating model convincing, an obvious next step is to move to short-run error-correction forms. But such models nest the long-run equilibrium relationship within a short-run model. In these dynamic models the a-theoretic time series dynamics usually provide most of the fit and forecasting ability, even at moderately long horizons. And even if there is agreement upon the long-run relationship, there will be multiple possible approaches for embedding this relationship within an error-correction form. Establishing the strengths of the long-run equilibrium model is a fundamental first step towards dynamic modeling.

Hence this paper avoids formal testing as much as possible, and instead follows two paths most models fear to tread. First, the static cointegrating model is used to forecast over long horizons, where "long" is well over five years. Second, in shorter-run forecasts the static model is compared to a dynamic pure time series model. Here the question is whether there are short horizons at which the static model clearly dominates.[3]

[3] This provides an informal definition of the long-run. The transition from short-run to long-run is the forecast horizon at which the static cointegrating model begins to dominate a pure-time series model.

1. The importance of scaling

The model presented below is simple, modifying standard approaches only in scaling money and GDP on a per household basis. This was suggested by Arthur Okun in a conference discussion of Goldfeld's (1973) exploration of scale variables in money regressions. In his exploration of the possible benefits of scaling Goldfeld had rejected a role for population in modeling total money.[4] But in his follow-up paper Goldfeld (1976) decided his tests were not valid and that theory implied a model employing money and income measured per-capita. Yet the issue of scaling and Okun's suggestion for scaling on a per household basis largely disappears from the literature after the 1970's.

1.1 Theoretical considerations for stability

Even in a world scaled up as a sum of representative agents, aggregate GDP growth is a function of both increases in individual incomes and increases in population. In itself, population growth merely replicates or re-scales the economy proportionately, so (barring externalities) this replication elasticity must be one. If the individual income elasticity does not also equal one, then a GDP elasticity will vary as population growth and individual income growth vary in their relative importance. So if population and individual income elasticities are unequal and their growth rates vary then we know a purely aggregate model must be unstable. Greene (1999) shows that for US data this instability will be large enough to matter if the individual income elasticity is less than 0.75. Results below imply a per household GDP elasticity well under this value.

[4] Meltzer (1963) also took an empirical approach to the issue, rejecting regressions of money per capita on income or wealth per capita because the R^2 was smaller than for non-scaled variables, and because there was more sub-period variation in coefficients for the per-capita models. In addition, he found that the "population elasticity" was not equal to one. These were static regressions in levels, but of course he estimated variances assuming stationary data.

1.2 Timing

If we grant the need for some sort of population scaling, then there is a second consideration. There is a lag between changes in the birth rate and changes in the number of managers of income and money. Even when looking at long-run trends, there will typically be about twenty years from a change in birth-driven population growth to the date at which this impacts the number of adults earning income and managing money holdings. So even if we are interested only in long-run relationships, something akin to Okun's scaling suggestion is essential for the timing of empirical data to correspond to theoretical notions.[5] Greene (2010) describes a number of alternative suggestions if data on the number of households is unavailable. These include using residential construction data, using measures of the adult population or the adult population less the number of married couples, or integrating up from birth, death and immigration rates.

1.3 An economically meaningless common trend

A third reason to use population or household scaling in estimating monetary models is that households will be cointegrated with households (or population with population), but such "self cointegration" is not very interesting. Denoting the log of households "h", the log of real money per capita "rmh", and the log of real GDP per capita "yh", then total real money is rmh +h, log real GDP is yh +h, and the standard aggregate model treats rmh +h as a function of yh +h. As changes in rmh and yh become smaller relative to changes in h, the model comes closer to treating households as a function of households.

[5] Ideally one would want a measure of independent economic units, and it may be possible to marginally improve upon Okun's suggestion.

For the USA from 1959 to 2008 the number of households (not logged) increased by a factor of about 2.28 and RGDP per household increased by a factor of about 2.11. So more than half the change in RGDP is due to mere re-scaling. Likewise, total real M1 increased by a factor of 2.76, but almost all of this is driven by replication, with real money per household only 1.21 times as large in 2008 as in 1959. Hence regressing total money on GDP (rmh +h upon yh +h) is more than halfway to modeling h as a function of h, treating "self-cointegration" as economically meaningful.[6]

2. M1, regulatory shadow-sweep accounting, and the long-run forecasting exercise

Starting in 1994 the Board of Governors essentially ended the requirement that banks report all depositor holdings of freely checkable accounts as such, allowing some to be reported as money market deposits with no reserve requirement. The limited discussion in the economics literature has referred to these as "retail-sweeps", but this is a misleading term.[7] Unlike true sweep accounts set up for the benefit of deposit owners, the balance "swept" into these shadow MMDA accounts does not pay money market interest to the depositor. The liquidity of the retail depositor is not affected. The retail checking account depositor writes unlimited checks regardless of the existence or non-existence of the shadow account. And the size and date of any "sweep transfer" is not reported to the depositor. So more transparent terms for this device are "shadow sweep", or "regulatory sweep". This is really a (de-)regulatory move in the same spirit

[6] Data on the number of US households equals the number of occupied housing units, as published by the US Census. Historical data is available as "Table HH-1, Households by Type: 1940 to Present" in the most recent Families and Living Arrangements/Historical Data section of the US Census Bureau website, http://www.census.gov/population/www/popdata.html. This source data is annual. In the empirical sections below these annual figures are interpolated to quarterly estimates.

[7] In the banking industry literature the device is sometimes referred to as a "reserve sweep".

as regulatory accounting standards which allow "off-balance sheet" activities which nonetheless affect the profitability of and risk borne by a financial institution.

An oddity of the new regulatory accounting rules was that they were proposed by an individual bank (O'Sullivan(1998, p88)). The banking committee of the US Senate and the Board of Governors treated this as a proprietary innovation. The Board further reasoned that if the innovating bank had been required to report the volume of shadow sweeps then this might have revealed details of the particular bank's activities to competitors (Melzer and Kohn in FMOC Minutes (Sept. 1995, p33)). Protecting the private innovation was more important than maintaining the reporting and measurement of freely checkable deposits.

Thus the Board of Governors decided not to require reporting of deposits in standard retail checking accounts which are treated for purposes of regulatory reserve-requirement accounting as money market deposits. Instead the Board proceeded to publish fictitious M1 data, with a growing portion of traditional checking balances hidden in the shadow sweeps. As of this date shadow-sweep accounting is estimated to hide nearly half of freely checkable deposits, so reported M1 is now about two-thirds of its true value from the depositor (and thus behavioral) perspective.

2.1 Adjusting the Fictitious Data

Once enough time passed to protect innovator privacy, the Board of Governors began to publish estimates of the volume of deposits held as M1 but not reported in the official data. To extend the M1 data beyond 1993, I take this data on estimated shadow sweeps and add it to M1 as reported by the Board. These adjusted figures are nearly identical to those developed by Cynamon, Dutkowsky, and Jones (2006), periodically updated and published as series M1RS at their website (http://www.sweepmeasures.com). But these are estimates. Banks are still not required to report the volume of freely checkable deposits hidden in the shadow MMDA accounts.

Anderson and Rasche (2001) report evidence that the accounting-sweep estimates are likely accurate. But unlike the remainder of measured M1, the monthly sweep data is so smooth as to contain no apparent seasonal components. Even if accurate on average, either there is no seasonal pattern or the estimates miss this detail. This may affect dynamic models employing M1 data adjusted for shadow sweeps, since seasonal components may shift sometime after 1993. But except for univariate (pure time series) models used as benchmarks, the focus here will be upon long-run static models.

2.2 The characteristics of aggregate M1, velocity and money per household

Figure 1 displays the trend characteristics of logged real M1. After 1982.3 there is a large increase in the trend slope. This change in trend comes a year after the shift in trend velocity, marked with the vertical line. Before 1982.3 there is some variation in trend, money grows at a steady rate 1959.1-73.1, and then declines a bit until the turning point in 1982.3. But these changes are dwarfed by the shift to a relatively constant increasing trend at 1982.3.

Figure 2 displays velocity in the sense of the ratio of real GDP to real M1. Velocity rises along a nearly linear path 1959.1-81.3. Then the trend ceases to be nearly linear. If one compares endpoints of the earlier to the later period (1981 versus 2008) there is little change in velocity, for a cumulatively flat trend. But along the way and for periods of as long as three years there are both positive and negative trends. So when comparing pre- and post-1981 time periods, velocity differs in both trend and variation in trend.

Figure 2 also displays real M1 per household. This largely mirrors the behavior of velocity, with two nuances. First, the turning point for trend money per household comes a bit later (1982.3) than for velocity (1981.3). This turning point for money per household is the same as for aggregate real M1 in Figure 1. Second, before 1981.3 money per household does not follow a linear trend as closely as does velocity. In particular, the trend of money per household shifts in 1971.3, before the larger shift in trend at 1982.3. So there is a bit more pre-1982 variation in the trends of real M1 per household than there is variation in trend velocity. But as with trend money, there is a large shift in trend velocity and real M1 per household after 1982.

3. Long-run static model forecasting

This section compares the long-run forecasting performance of two long-run models estimated via Engle-Granger static levels, one written in standard aggregate terms and the other with money and RGDP scaled per household. The first forecasting exercise estimates through 1981.3, while the second estimates the models through 1987. In the above section we saw that whether viewed through the lens of aggregate money, velocity, or money per household the 1959.1-81.3 data excludes a large shift in trend. Hence I characterize long-run forecasts based upon estimation within this period as "most ambitious". The terminal estimation date for the second forecasting exercise is chosen to include five years of data after the shift in trend money (1959-87), and so is less ambitious.

3.1 A most ambitious exercise: Forecasts through 2008 based upon estimation 1959-1981

Here the simple static cointegration (levels) models are estimated through the date of peak velocity 1959.1-1981.3. The coefficient estimates are then used to forecast the remaining sample through 2008.4, taking the interest rate and income at actual values (conditional forecasts). The two models differ only in whether the measure of money and income is scaled. The first model regresses (logged) aggregate real M1 upon a constant (rm), real (chain weighted) GDP (y), and an interest rate (r), the 10-year Treasury (constant maturity) yield. Nominal money is deflated using the chain-weighted GDP deflator. The second model employs the same variables, but money and GDP are scaled per household (rmh and yh). The coefficient estimates for the aggregate model and the per-household model are (respectively)

$$r\hat{m} = 4.103 + 0.333y - 0.132r \quad \text{and} \tag{1}$$

$$r\hat{m}h = 7.748 + 0.204yh - 0.356r \tag{2}$$

and forecasts in Figures 3 and 4 use these estimated values.

In Figure 3 the forecasts of the aggregate model of Equation 1 have been converted to per household terms by subtracting the log of households. For the first couple years after the turning point, the aggregate model appears to track the reversal of trend money (per household). But by the late 1980's the forecast values from the aggregate model are consistently drifting farther away from actual values. The aggregate model is a limited success in the sense that it does not predict a continuation of the pre-1981 downward trend. But it is unable to track the large increase in real money per household over the subsequent decades.

Figure 3 also shows the forecasts of the per-household model of Equation 2. Besides tracking the reversal in trend, the model predicts the level remarkably well, although with smaller errors before the introduction of shadow sweeps. Before the introduction of shadow sweep regulatory accounting, the predicted values cross actual values six times. After the introduction of shadow sweeps (after the transition to a more crudely measured M1) the predicted values do not cross actual values until more than ten years have elapsed, at the end of 2007.

Figure 4 shows the same data and results, but this time the number of (logged) households is added to the forecasts of the per household model, converting them to aggregate terms. As with the per household data, the turning point in the trend of aggregate money comes a few quarters after peak velocity (1981.3). But from this aggregate money perspective, the aggregate model predicts a trend much the same as trend money over the estimation period. In contrast, the converted forecasts of the per household model catch the change of slope in aggregate money.

3.2 A less ambitious exercise: Forecasts based upon estimation 1959-1987

The estimation period used above excluded the shift in money and velocity that occurs after 1981. Including some post-81 data on money will markedly increase the variance of money trends used to obtain model estimates. In fact, the most positive 5-year trend slope of real M1 per household begins in 1982.3, a year after the shift in velocity. So here I extend the estimation period through 1987. This estimation period includes both the most positive and most negative 5-year trends of aggregate real money and of money scaled per household over the entire period 1959-2008.

For 1959.1-1987.4 coefficient estimates for the aggregate model and the per-household model are (respectively)

$$r\hat{m} = 2.592 + 0.545y - 0.258r \quad \text{and} \tag{3}$$

$$r\hat{m}h = 7.406 + 0.237yh - 0.368r \quad . \tag{4}$$

Notice the ratios of the estimated coefficients of Equation 3 and those of Equation 1 are respectively (for the constant, y and r) 0.63, 1.64 and 1.95 This is a substantial shift in the estimated coefficients of the aggregate model. For the per household model the corresponding ratios are much closer to one, the ratio of estimated coefficients of Equation 4 to those of Equation 2 equaling respectively 0.96, 1.16 and 1.03 . In the economic sense the per household model estimates are relatively constant.

Figure 5 displays the fitted values and forecasts from these models. Here including the steep change in trend money 1982-87 does help the aggregate model, in the sense that forecast errors do not increase over time as much as in Figures 3 and 4. But the forecast errors of the aggregate model are still much larger than those of the per household model, and the aggregate model forecasts still drift substantially away from trend money.

4. Data characteristics and the performance of alternative forecast estimators

This section examines three questions. First, are the per household variables I(1), as expected for their aggregate counterparts? Second, do cointegration tests reject non-cointegration among the scaled variables? Third, for the "ambitious" forecasting exercise (as above), do alternative long-run estimators improve upon Engle-Granger OLS?

4.1 Integration and cointegration.

Table 1 shows the results of ADF tests for the integration properties of the data used in the scaled (per household) model. Tests are conducted with and without a constant in the test regression, and also including a constant and time trend, corresponding to three separate test assumptions. In addition, test regression lags are selected to minimize both the Schwarz and Akaike criteria, for a total of six tests for the null of a unit root applied to each variable.

For each levels variable Table 1 reports the minimum probability value across the test assumptions. For the differenced data the maximum probability value across tests is reported. Thus for the levels variables the reported test is that which leans towards rejection of a unit root (actual test size is greater than this reported p-value), while the reported results for the differenced data is biased towards accepting the unit root (actual test size is less than the reported p-value). Despite this bias, a unit root is accepted for the variables in levels (with a smallest p-value of 0.24), while a unit root is rejected for the differenced variables (with a largest p-value of 0.002). The results of Table 1 imply the scaled variables have the properties usually encountered in monetary and income aggregates.

Table 2 displays Johansen cointegration test results for a null of no cointegrating relationships (versus more than zero) among logged real M1 per household (rmh), logged real GDP per household (yh) and logged yields on Treasuries of 10-year maturity. Under the null the model is a differenced VAR, so lag selection is done in this differenced VAR (no levels or error-correction included). The AIC selects three lags, and the SIC selects one lag, so results are displayed for both truncations. Under both lag-selection criteria, Table 2 shows that a null of no cointegrating relationships is rejected, with (nominal) probability values of under one percent.

4.2 Sample information as experimental design

In long-run estimation we are trying to distinguish meaningful from spurious common trends. If the data came from controlled experiments with income and the interest rate as the controlled variables, then a well designed experiment would impose substantial independent variation in the trends of these variables (r and yh). In addition, a well designed experiment would impose sustained shifts in trends, so the variation in trend money would be dominated by the long-run relationship rather than by short-run dynamics. From this perspective history has given us a poorly designed experiment. But it turns out that although far from ideal, the pre-1982 movements in income and the interest rate are closer to a well designed experiment than the post-1982 data.

Table 3 displays sample variances and correlation coefficients for 5-year trends (differences) of the (logged) interest rate and income. I chose five years because as displayed in Figure 3, fitted values seldom diverge from actual values for more than five years.[8] Reading down the first column for Δyh, the variance of real GDP per household Δyh in the earlier sample (1959.1-81.3) is 3.33 $*10^{-3}$, which is about one-third larger than in the sample as a whole (2.43 $*10^{-3}$). These variances for Δyh are about an order of magnitude smaller than the sample variances for trends in the interest rate.

This would be a reasonable experimental design if the cointegrating coefficient on yh was an order of magnitude larger than that for the interest rate. But in most empirical studies and theoretical models these coefficients are well within the interval [-1, 2]. And in Equation 4 the absolute values of the estimates differ by less than 0.2. So historically the variation in yh is much too small for a well designed experiment.

The entire sample 1959-2008 presents us with an additional element of a poorly designed experiment. Ideally, there would be no correlation between the trends of yh and the trends of r. But as displayed in the first row of Table 3, the correlation is about -0.2. So 1959-2008 there is too little variation in income, and what little variation exists is not independent of the interest rate.

[8] Results for relative variances and correlation are similar for first through 40'th differences.

The poor experimental design is improved if we look only at the earlier period 1959.1-1981.3. The last row of Table 3 displays the trend variances and correlation as calculated only over this period. The sample variance of Δyh of $3.33*10^{-3}$ is more that a third greater than for the sample as a whole. More importantly, the correlation of Δyh and Δr falls to nearly zero (0.017). So in the earlier portion of the data the variation in the trends of RGDP per household and of the interest rate are more informative.

This has two sorts of implications. Starting estimation at a later date will omit the portion of data containing the most information. So unless there is substantial instability in the cointegrating model relationship, forecasts based upon moving-window estimation will have larger errors than forecasts based upon including the earlier sample. Second, if the recovery from the 2008 recession is particularly slow or stalls, or if the recovery is particularly steep and sustained, then this will represent the first sustained shock to trend yh in about three decades. So updating this study to include all of the post-2008 recovery may be of particular interest.

4.3 Long-run forecasts from alternative estimators, DOLS and FM-OLS.

This section compares the forecast performance of alternative estimators of the cointegrating relationship in the "most ambitious" and "less ambitious" forecasting exercises discussed above.[9] Both FM-OLS (as in Phillips and Hansen (1990)) and DOLS (Saikkonen (1982)) estimators are compared to the static Engle-Granger approach above. For DOLS and FM-OLS methods a variety of estimators are possible depending upon lag selection criteria and/or choice of kernel and kernel bandwidth. Some confusion of terminology is possible here. Both methods employ dynamic models. But the dynamic element is used as a means to produce

[9] Results for CCR as in Park (1992) were almost identical to that for FM-OLS, with usually slightly larger forecast RMSE.

alternative estimators of the coefficients in the static cointegrating model. Thus the form of the model used in these alternative forecasts is the same as above in Section 3, only the form of the estimator for the $\hat{b}$'s varies across the methods.

For DOLS estimation I started with the maximum symmetric number of leads and lags of differenced variables possible. In both estimation samples (terminating in 1981 or 1987) the SIC increased with any simplification.[10] In FM-OLS for estimation of the variance-covariance matrix I used the most commonly applied Bartlett kernel. In choosing the kernel bandwidth I employed the "fixed" Newey-West (1994) approach, which sets bandwidth on the basis of sample size.[11]

Table 4 displays the forecast RMSE of the competing static models. The first column shows results for the most ambitious forecasting exercise, including the simple Engle-Granger approach of Figures 4 and 5. The forecast RMSE employing DOLS estimation is 0.107, about a third larger than the RMSE (0.070) of the Engle-Granger estimator. The forecast RMSE for FM-OLS (0.047) is less than half that from DOLS, and about a third less than for Engle-Granger estimation.

The last column of Table 4 estimates models for the less ambitious exercise, extending the estimation to include five years after the major turn in trend money. The difference between the models narrows somewhat, but the DOLS model forecast RMSE is about thirty percent larger than the Engle-Granger model RMSE, and likewise the difference between Engle-Granger and FM-OLS is about thirty percent. Since DOLS does not yield an improvement over the simple Engle-Granger approach, I drop it from consideration in the remainder of this paper.

[10] Forecasts from the model estimated with 2, 4, or 8 leads and lags resulted in larger RMSE.
[11] The term "fixed" is used because other methods depend upon characteristics of the data that can differ with fixed sample size.

5. Shorter-run forecasting: Static cointegrating versus dynamic pure time series models

In forecast competitions pure time series models often out-perform economic models. When an economic model performs relatively well, the economic model usually nests within it pure time-series components. With integrated variables pure time series is inherently short run (differenced ADL), while the cointegration model is inherently long-run. Nonetheless it is possible to make an interesting short-run comparison of the two approaches. This section considers the RMSE of a series of 1-step-ahead forecasts, where "step-ahead" references the updating of coefficient estimates as the sample terminus and the forecast date is extended. But it also considers forecasts a K'th step-ahead, through K = 20 quarters.

This allows us to ask three related questions. Is there a time horizon at which the static cointegrating relationship forecasts with lower RMSE than a differenced time series model? If so, at what horizon does the static long-run model begin to beat the dynamic short-run model? Third, is the difference in performance large enough to matter? Put differently, how many quarters is clearly the short-run, where pure-time series (differenced) ADL models dominate, and how many quarters ahead is the long-run, where the cointegration model clearly dominates?

I refer to time series models (plural) because each time the estimation sample is extended by an additional quarter, I re-specify the lags retained to minimize the SIC. The forecast from the quarterly differenced model for the level a K'th step ahead is generated by iterating the forecast difference forward K-1 times ("dynamic forecasts") and then summing the total forecast differences. In the static cointegration model forecasts the estimation is updated at each step, which for FM-OLS implies the bandwidth is increasing as the sample is extended.

Figure 6 displays the forecast RMSE from the competing models as a function of the forecast horizon (K). First, note that the static cointegrating models begin to have a smaller forecast RMSE at a time horizon of 6 quarters. Thus the long-run models begin to contain superior information at a horizon of about a year and a half, in the sense that the static cointegrating model outperforms the short-run time-series model at that point and beyond.

Second, note that the gap between the time-series and cointegration models increases steeply with the horizon. By a horizon of 12 quarters the RMSE of the pure time series model is twice as large as the others, and by 20 quarters is three times as large. And the cointegrating model's forecast RMSE is remarkably constant, there being no obvious slope to the function as the forecast time horizon is extended from 1-quarter-ahead through 20-quarters-ahead. Soon after two years we are clearly “in the long-run”, in the sense that the long-run model strongly dominates the short-run time series model.

Finally, note the performance of simple Engle-Granger estimation is nearly identical to that of FM-OLS. With this data and forecast time horizons of five years or less, there is little or no gain to adopting the more complex FM-OLS estimator. So in the remainder of this paper I apply FM-OLS only when valid standard errors are of interest.

6. Potential for inflation modeling

Cointegration of (logged) real M1 per household, real GDP per household and the 10-year treasury yield implies cointegration of the latter two variables with the price-level and nominal M1 per household (mh). The real money specification imposes a unitary weight on the price-level, implying the cointegrating model

$$p = mh - (b_0 + b_1 yh + b_2 r) + \varepsilon. \qquad (5)$$

Using $\hat{rmh}$ from the cointegrating model employed in previous sections to generate $\hat{p} = mh - \hat{rmh}$ is merely a restriction of the unitary weight price model of Equation 5. A distinct model will estimate

$$p = b_3 mh - (b_0 + b_1 yh + b_2 r) + \varepsilon, \qquad (6)$$

without the constraint $b_3 = 1$. This section considers the sense in which the long-run money model has potential for use within a price-level and inflation model, rearranged into the form of Equations 5 or 6.

6.1 Cointegration, estimates and fitted values from long-run models

Table 5 displays the results of Johansen tests for cointegration among p, mh, yh, r and a constant, respectively the GDP deflator, nominal m1 per household, real GDP per household and the 10-year US Treasury yield 1959.1 -2008.4. Non-cointegration is rejected with p-values of 0.0001 or less. This test does not impose the unitary nominal money elasticity implicit in the models of real M1.[12] Estimating the cointegration relationship via FM-OLS, the estimates with and without imposing a unitary elasticity are presented in Table 6.

In addition to FM-OLS estimates, Table 6 presents simple Engle-Granger estimation results. Three characteristics stand out. First, the unrestricted estimates for the coefficient on nominal money per household (mh) are close to one, differing by 0.017 in FM-OLS or by 0.035 in simple Engle-Granger estimation. Second, for FM-OLS the unitary elasticity restriction is within half a standard error (0.041) of the unrestricted estimate of 1.017. Third, when the unitary elasticity is imposed then FM-OLS estimates are close to the simpler Engle-Granger estimates.

[12] Imposing this restriction would merely repeat the results of Table 2.

Figure 7 displays actual and fitted levels of the GDP deflator from the restricted Engle-Granger regression (Equation 6), and also displays forecasts as implied by estimates from the sample 1959.1-1981.3, equivalent to the "ambitious" money forecasts of Figure 3. These forecasts of the price-level track very well, especially considering the long forecast horizon.

But consider (by eye) the slope of the values as fitted over the whole sample displayed in Figure 7. Over periods of a year or less the slope of fitted values is often steeply positive or negative, implying inflation or deflation of magnitudes far from actual values. If the cointegrating level estimates are to be substantially useful in inflation forecasts, then we must be interested in inflation over longer time horizons.

6.2 K'th step-ahead forecasts of inflation at horizons of one to ten years

As in the forecasting exercise of Figure 6, here consider K'th step-ahead forecasts based upon the cointegrating model of the price-level (Equation 5) and a differenced ADL model with lag selection criteria re-applied at each increment of the terminal date of estimation. And as in the forecasts of Figure 4, forecasts begin in 1980, thus always including the period 1959.1-79.4, which includes most of any variation in income trends independent of interest rate trends. For the cointegrating model the level forecasts K steps ahead ($\hat{p}_{t+K}$) are converted to a forecast of the implied annualized inflation rate as

$$\hat{\pi}_K^{LR} = \exp[(\hat{p}_{t+K} - p_t)(4/K)] - 1. \qquad (7)$$

The ADL model is estimated in logged quarterly differences, with forecasts iterated ahead (dynamic forecasts) producing a forecast of the cumulative change from time t to t+K ($\hat{\Delta}_K p_{t+K}$), which is then transformed to an annualized inflation rate forecast as

$$\hat{\pi}_K^{ADL} = \exp[(\hat{\Delta}_K p_{t+K})(4/K)] - 1. \qquad (8)$$

Figure 8 displays the inflation forecast RMSE for each model as a function of the forecast horizon (K), for K = 4 through 40 quarters ahead. The RMSE of the time series model is initially 1%, increasing slowly and almost linearly to 2% for 10-year ahead forecasts. The forecast RMSE of the static cointegrating model is initially 4%, or four times that of the time-series model. The static model forecast RMSE does not fall below that of the time-series model until the 14-quarter horizon. The difference between the two models increases slowly, so at a 5-year horizon the difference is only half a percentage point. For a full percentage point difference the forecast horizon has to be seven years or more. If the "long-run" is the horizon at which the static cointegration model clearly outperforms the ADL model, then for inflation the long-run is very long indeed.

I interpret these results as implying that the usefulness of the cointegrating model in inflation forecasting is marginal, or certainly less evident than performance in forecasting money itself. If one compares the actual path of money per household in Figure 5 to the path of the price level in Figure 7, the reason for the differential performance can be seen in the differential smoothness of the levels data. Post 1980 a linear trend will fit the price-level data much more tightly than a linear trend will fit money per household. Hence even at long time horizons a joint time series/cointegration model (i.e. an error-correction model) would place substantial weight upon past trends. A long-run relationship is less important in inflation forecasting than it is in forecasting of money (M1).

7. Conclusion

This paper has proceeded on the premise that most readers are skeptical of money modeling, hence it will take strong results to budge prior beliefs. Two sorts of ambitious demonstrations are presented. In the first demonstration, a simple static cointegration model is estimated over a time period in which velocity and money per household trend downwards, with the coefficient estimates used to forecast from 1982-2008, a forecast period of more than 25 years in which trend velocity and money have reversed. The model tracks remarkably well.

In the second demonstration the model forecast RMSEs a K'th step ahead are compared to those from a pure time series model. In most such forecasting contests the static cointegration model would be nested within a dynamic form. But here the cointegration model of money competes with the time series benchmark without the benefit of added dynamic components. Yet the static per household cointegration model beats the time series benchmark at a forecast horizon of only six quarters, with the difference in forecast RMSE increasing steeply as the horizon increases. At 20 quarters the forecast RMSEs differ by a factor of three.

Along the way properties of the data and alternative cointegrating estimators are explored. The simple theoretical ideas presented in Section 1 imply it is important to distinguish trends in income per household from other trends. But most of the variation in trend income occurs before 1982, and the post-1982 variations in trend income are more highly correlated with interest rate trends. So moving-window methods which exclude the earlier data are unlikely to perform well. And once a model successfully forecasts the large shift in trend money occurring in 1982, there is relatively less for the model to forecast (the great moderation contains little information). An important out-of-sample exercise will be to check the performance of the model after the recovery from the 2008 recession is well established, as this period will likely include changes in trend income (and trend money) not seen in 30 years.

In the ambitious forecasts (estimation terminating in 1981.3 or 1987.4), the FM-OLS estimator produces forecasts with substantially smaller RMSEs than the simpler Engle-Granger estimator. But this advantage disappears in the more traditional K'th step-ahead forecasting exercise. Forecasts based upon a DOLS estimator perform poorly, with lead-lag selection criteria implying retention of so many lags as to substantially reduce the degrees of freedom for estimation.

The money model can be rearranged to model the price level, and so a brief exploration of potential in forecasting inflation is presented. The relative smoothness of inflation implies pure time series properties will dominate over longer horizons than when modeling money. Given the pervasive finding in literature that pure time-series models dominate other models of inflation (as surveyed in Stock and Watson (2009)), some readers will take the performance of the price-level cointegrating model to be impressive. But the gains over a pure time series model are not substantial until the horizon is about five years, at which point the forecast RMSE amounts to a bit more than half a point. I take this to imply that exploration of any potential for inflation modeling will require a move to forms which nest cointegrating and dynamic time series models. This is in contrast to the ability of the long-run static model to stand on its own when modeling real M1.

References

Anderson, R., and Rasche, R. (2001). "Retail Sweep Programs and Bank Reserves, 1994-1999", *Federal Reserve Bank of St. Louis Review*, 83, (1), 51-72

Cynamon, B., Dutkowsky, D., and Jones, B. (2006). "Redefining the monetary aggregates: A clean sweep", *Eastern Economic Journal* 32 (4), 661-673

FMOC (1995), "Transcript, Meeting of the Federal Open Market Committee, September 26, 1995", Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/monetarypolicy/fomchistorical1995.htm

Goldfeld, S. (1973). "The Demand for Money Revisited", *Brookings Papers on Economic Activity*, 3, 577-638

Goldfeld, S. (1976). "The Case of the Missing Money", *Brookings Papers on Economic Activity*, 76, (3), 683-730

Greene, C. (1999). "On the Impossibility of a Stable and Low GDP Elasticity of Money Demand: The Arithmetic of Aggregation, Replication and Income Growth", *Applied Economics*, 31, (9), 1119-27

Greene, C. (2010). "Smooth-Adjustment Econometrics and Inventory-Theoretic Money Management", *Journal of Economic Dynamics and Control*, 34, (6), 1031-47

Hess, G., Jones, C. and Porter, R. (1998), "The Predictive Failure of the Baba, Hendry and Starr Model of M1", *Journal of Economics and Business*, 50, (6), 477-507

O'Sullivan, O. (1998). "Counting cash in the dark", *ABA Banking Journal*, 90, (2), 86-92

Phillips, P. (1990). "Statistical Inference in Instrumental Variables Regression with I(1) Processes", *Review of Economic Studies*, 57, (1), 99-125

Saikkonen, P. (1991). "Asymptotically Efficient Estimation of Cointegration Regressions", *Econometric Theory*, 7, (1), 1-21

Stock, J. and Watson, M. (2009). "Phillips Curve Inflation Forecasts", In Fuhrer, J. (ed), *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective*, MIT Press, Cambridge, 101-86

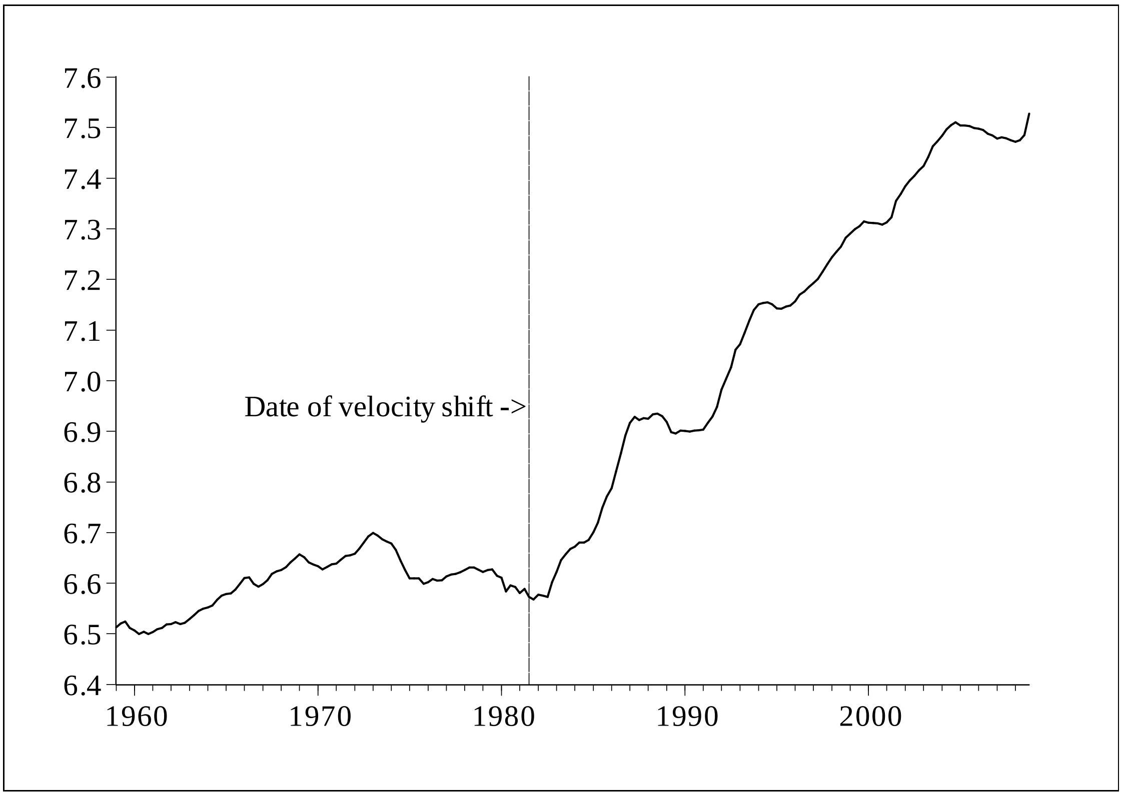

Figure 1. Shifting trends in logged real M1 1959-2008

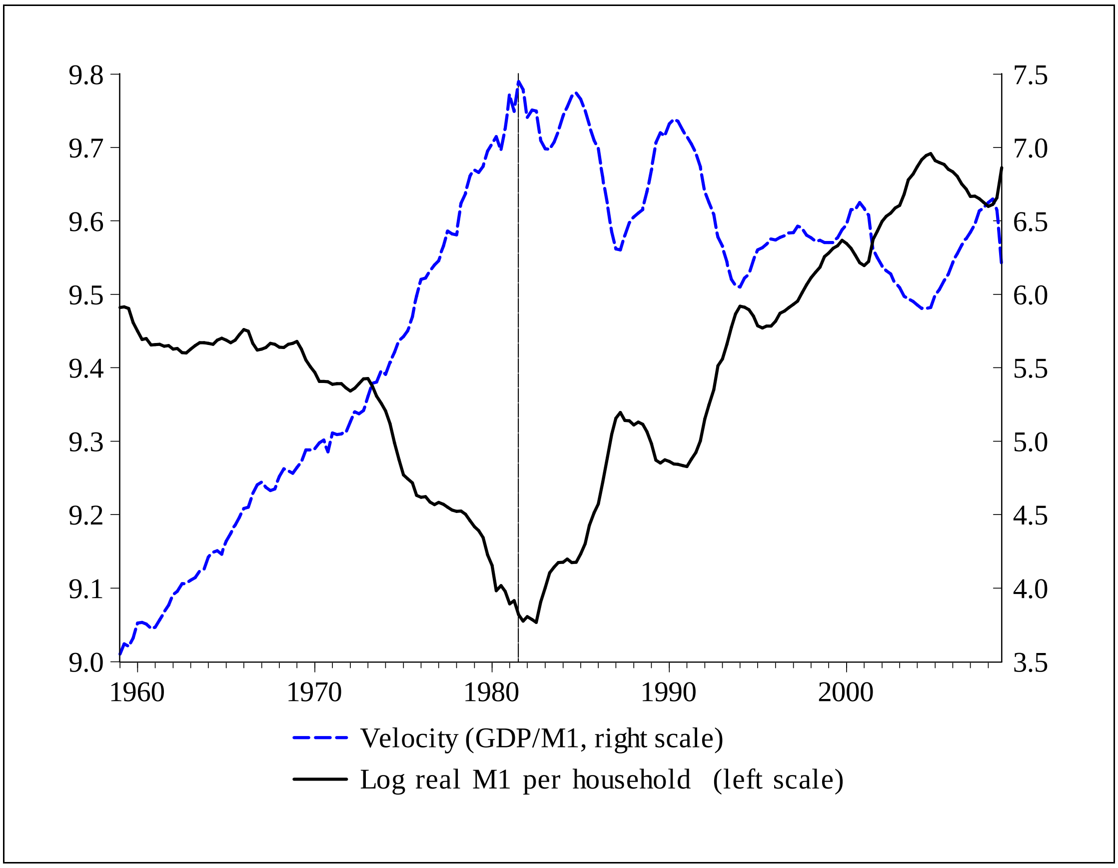

Figure 2. Shifting trends in velocity and real M1 per household.

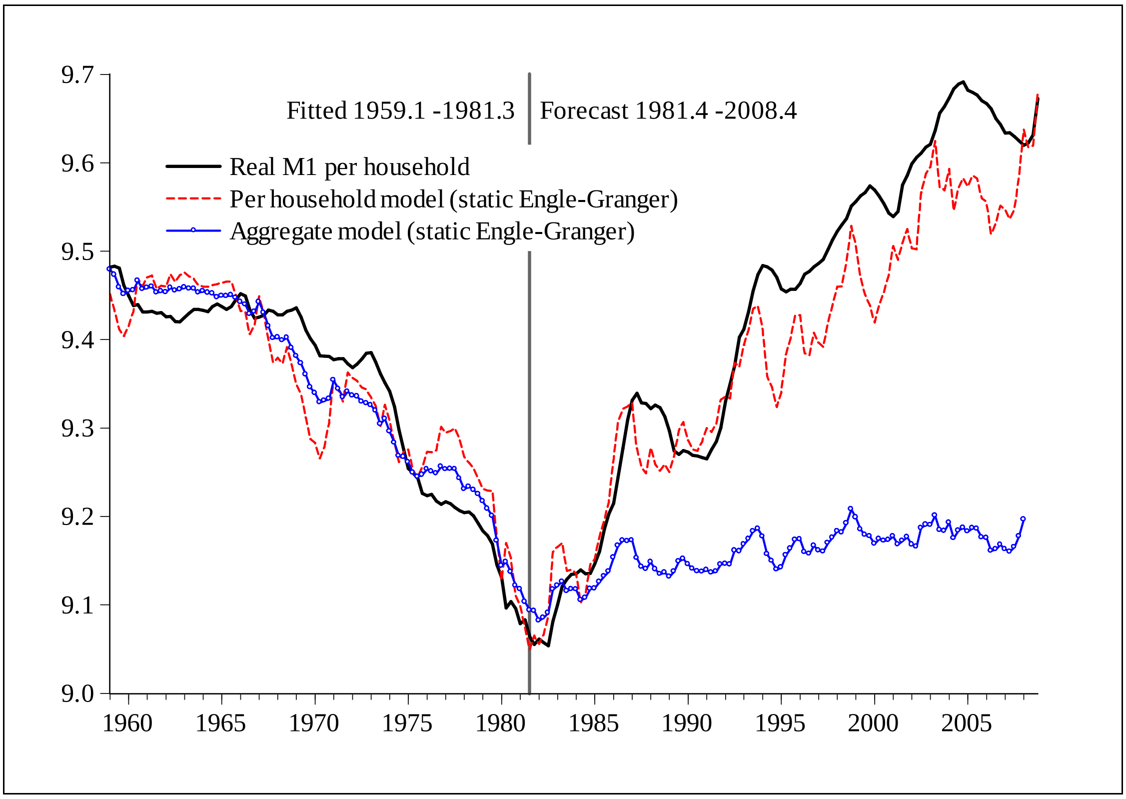

Figure 3. A most ambitious long-run forecasting exercise. Estimation sample terminates at shift in trend velocity (and just before shift in trend money). Static Engle-Granger model of the form of Equations 1 or 2 is estimated 1959.1-1981.3, then projected forward using actual values of income and the interest rate. Aggregate model fitted and projected values are transformed to per household values. Model forecasts treat income and interest rate as known.

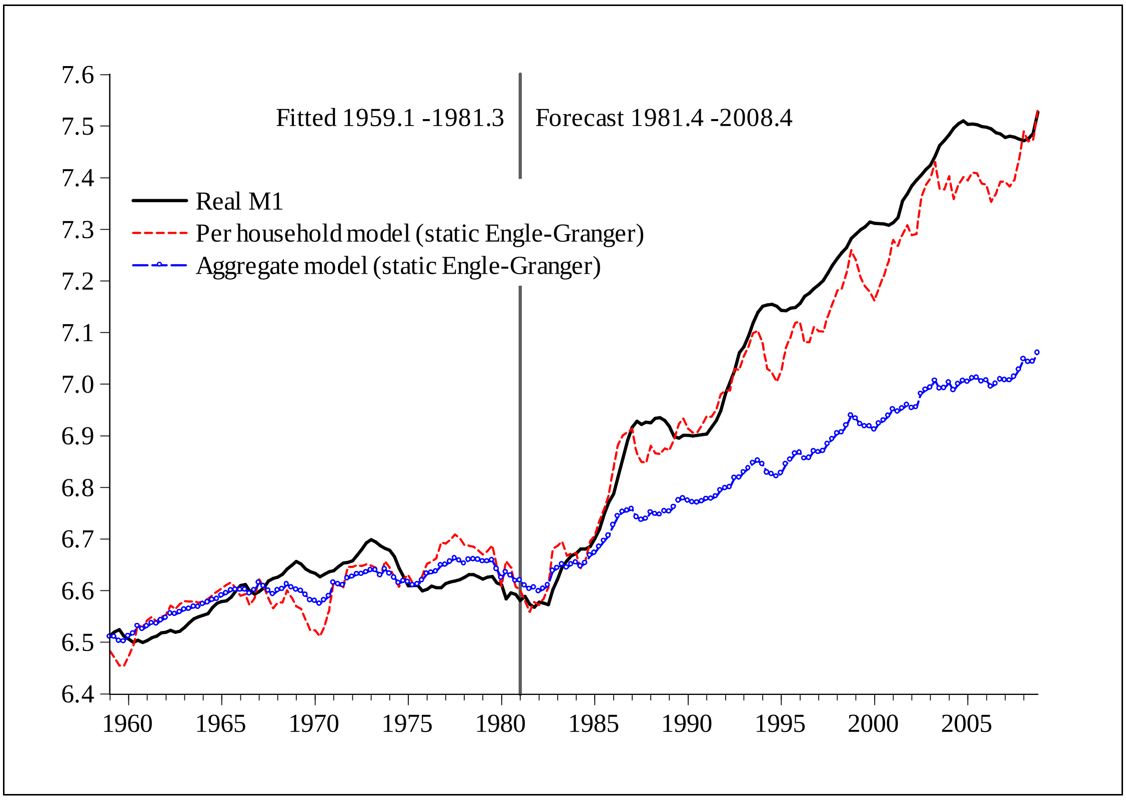

Figure 4. The most ambitious forecasting exercise of Figure 3 displayed in aggregate terms. Per household model fitted and projected values are transformed to aggregate values.

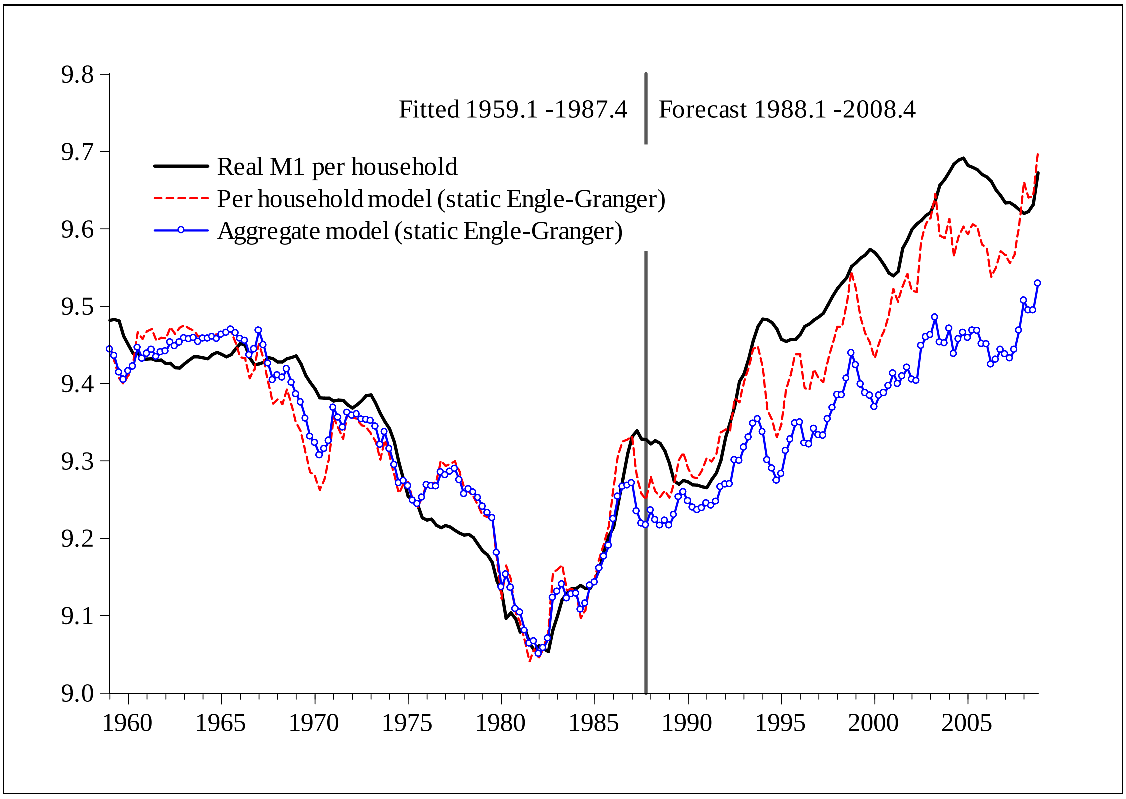

Figure 5. A less ambitious long run forecasting exercise. Estimation sample includes 5 years after money trend shift. Static Engle-Granger model of the form of Equations 3 or 4 is estimated 1959.1-1987.4, then projected forward using actual values of income and the interest rate. Aggregate model fitted and projected values are transformed to per household values.

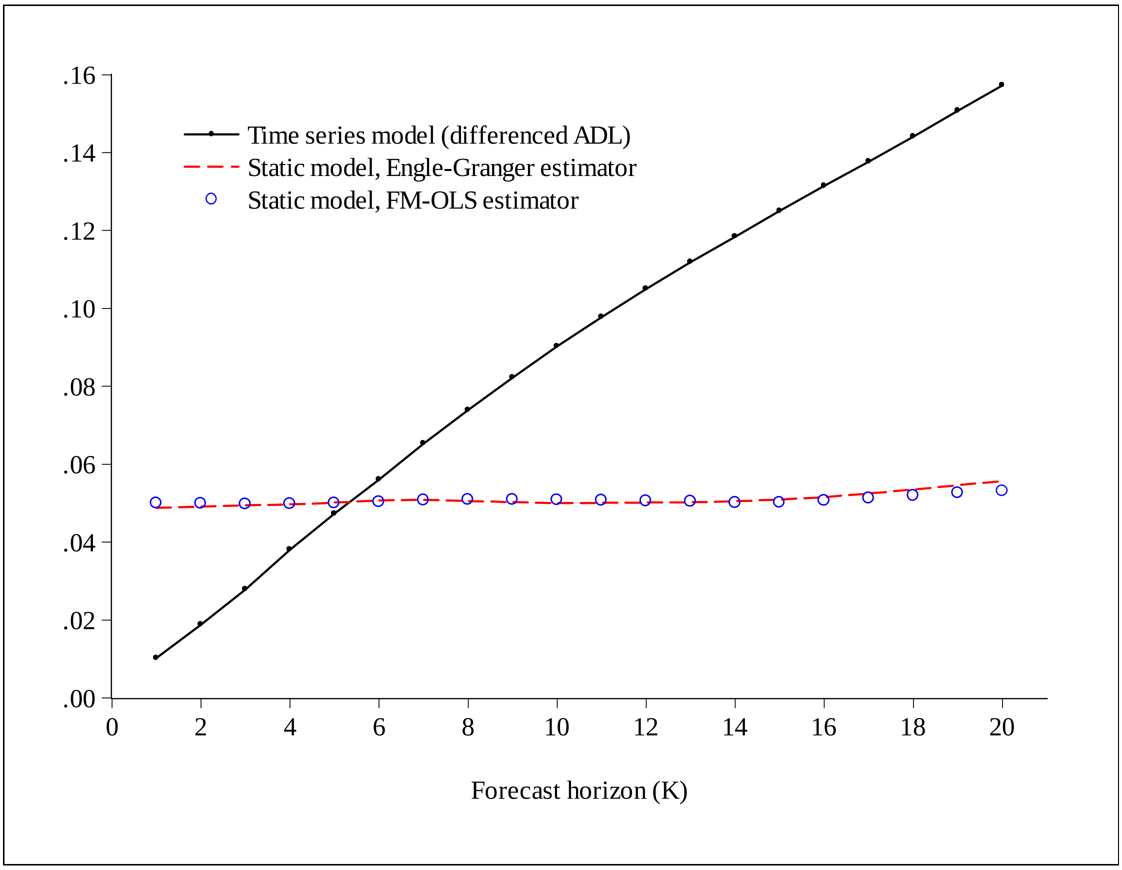

Figure 6. Forecast RMSE 1980-2008 of real money per household as function of forecast horizon (K). Models are initially estimated 1959.1-1979.4, so forecasts begin with 1979.4 +K. In projections static models employ actual values of income and the interest rate (yh and r). Lags in the time-series model were truncated to minimize the SIC, with the unconstrained model including eight lags. Lag length for this model was re-specified as the estimation sample was extended quarter-by-quarter, with a maximum of 6 lags retained. Static cointegration model forecasts treat income and interest rate (yh and r) as known.

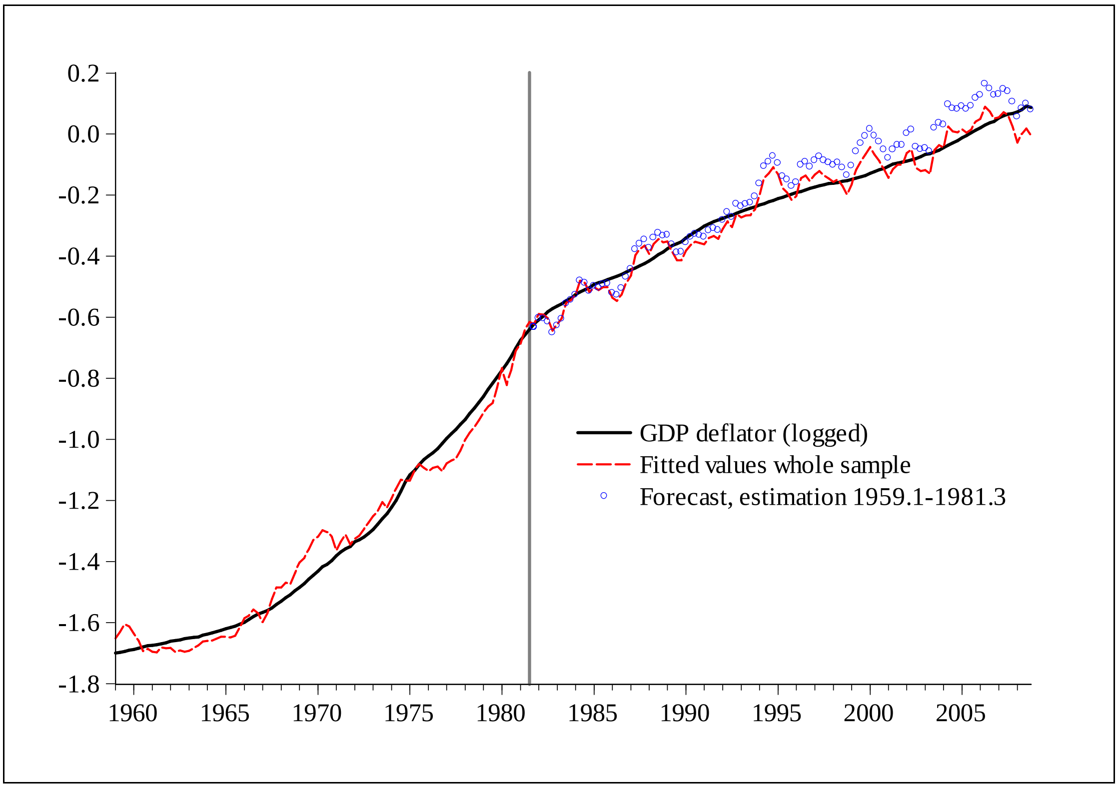

Figure 7. Price-level fitted and ambitious forecast values from inverted money demand model (Equation 5). Forecast values from estimating coefficients 1959.1-1981.3 employ actual values of nominal money, income and the interest rate (mh, yh, r).

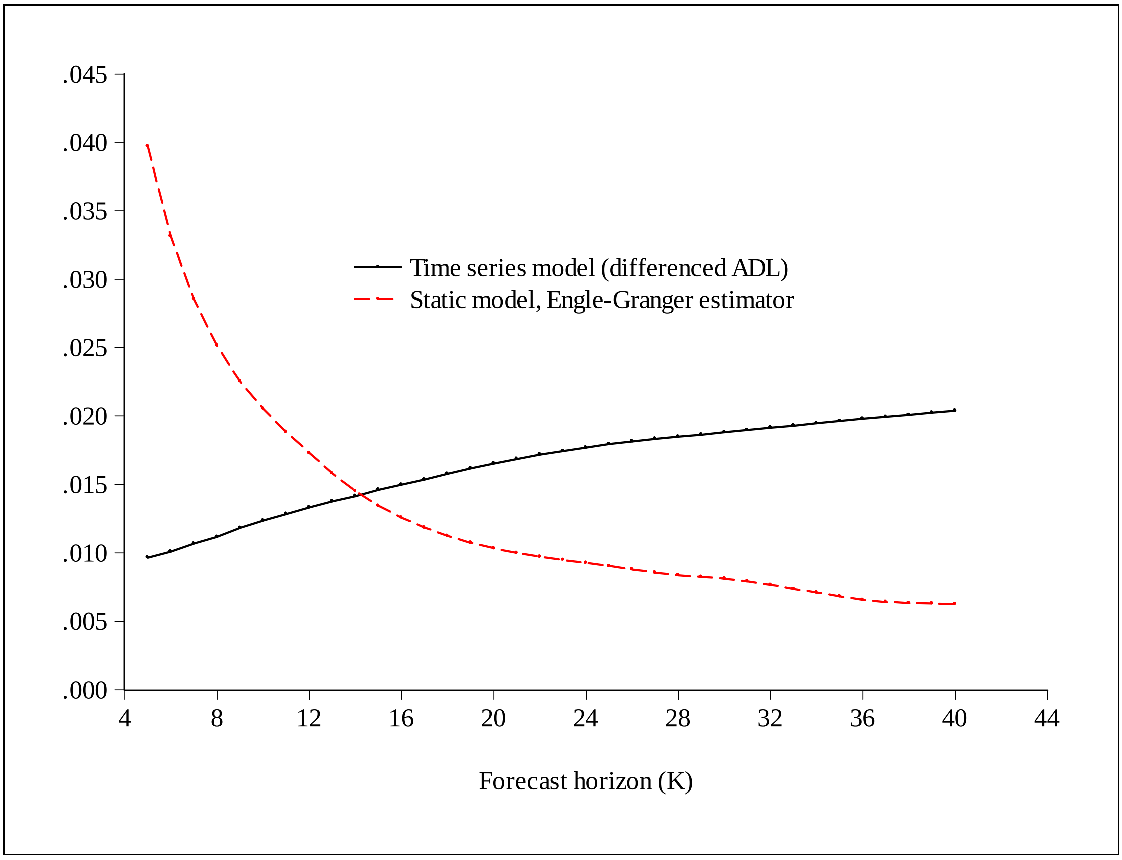

Figure 8. Forecast RMSE 1980-2008 as function of forecast horizon (K). K'th quarter-ahead forecasts of annualized inflation from static cointegrating (inverted money demand) and pure time series models. Lags in the time-series model were truncated to minimize the SIC, with the unconstrained model including eight lags. Lags retained always equaled one. A single-lag time-series model imposing a unit root resulted in slightly larger RMSEs. Static cointegration model forecasts treat mh, yh and r as known.

Table 1
Integration of per-household model variables:
ADF tests across assumptions and lag criteria.

Levels variables	Min p-value across tests
mh (Log real M1 per household)	0.86
yh (Log real GDP per household)	0.24
r (Log 10-year Treasury yield)	0.50

Differenced Variables	Max p-value across tests
Δmh	0.002
Δyh	less than 0.001
Δr	less than 0.001

Notes: Before differencing and lag-truncation data is quarterly 1959.1-2008.4. Test regressions are constructed with both min AIC and min SIC lag criteria, and with and without a constant or constant and trend. Lags retained varied between eight and zero. Results are as calculated by Eviews, which employs probability values from MacKinnon (1996).

Table 2
Johansen Cointegration Tests
Per household model variables (rmh, yh, r, constant) 1959.1-2008.4

VAR lag criteria	Lags retained	Trace statistic (No cointegrating vectors)	Probability value
AIC	3	42.64	0.007
SIC	1	50.13	less than 0.001

Notes: Probability values are less than 10% for three through six lags. The aic and sic lag selection criteria were applied to a differenced VAR, which is the correct model under the null of no cointegration. Results are calculated using Eviews, which takes p-values from MacKinnon, Haug and Michelis (1999).

Table 3
Ex-post experimental design:
5-year trends in the interest rate and real gdp per household

Sample	Sample variance $\Delta_5 yh$	Sample variance $\Delta_5 r$	Correlation $(\Delta_5 yh, \Delta_5 r)$
1959.1-2008.4	$2.43 * 10^{-3}$	$8.00 * 10^{-2}$	-0.209
1959.1-1981.3	$3.33 * 10^{-3}$	$2.64 * 10^{-2}$	0.017

Notes: $\Delta_5 x$ is the twenty-quarter difference $x_t - x_{t-20}$. Correlation and relative variances of Δyh and Δr are similar for one- through 40-quarter differences.

Table 4
Forecast RMSE for real M1 per household in ambitious and less ambitious exercise: Alternative cointegrating estimators.

Estimation Method	Estimate 1959.1 -1981.3 Forecast 1981.4 -2008.4	Estimate 1959.1 -1987.4 Forecast 1988.1 -2008.4
Engle-Granger	0.0701	0.0662
DOLS	0.1073	0.0926
FM-OLS	0.0466	0.0464

Table 5
Price-level model from inverted money demand, without imposing unitary price-elasticity
Johansen cointegration tests (p, nominal mh, yh, r, constant) 1959.1-2008.4

VAR lag criteria	Lags retained	Trace statistic (No cointegrating vectors)	Probability value
AIC	7	77.60	0.0001
SIC	1	97.67	< 0.0001

Notes: The aic and sic lag selection criteria were applied to a differenced VAR, which is the correct model under the null of no cointegration. Results are calculated using Eviews, which takes p-values from MacKinnon, Haug and Michelis (1999).

Table 6
Price-level cointegrating estimates, unrestricted and restricted 1959.1-2008.4

$$p = \hat{b}_0 + \hat{b}_1 mh + \hat{b}_2 yh + \hat{b}_3 r$$

Estimation	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$
FM-OLS	-5.812	1.017	-0.406	0.390
(SE)	(1.075)	(0.041)	(0.126)	(0.017)
FM-OLS	-6.305	impose b_1 =1	-0.350	0.399
(SE)	(0.327)		(0.029)	(0.017)
Engle-Granger	-5.555	1.035	-0.441	0.382
Engle-Granger	-6.421	impose b_1 =1	-0.337	0.382

Notes: As above, FM-OLS employs Bartlett kernel with Newey-West “fixed” bandwidth based upon sample size, here equal to 5 quarters).