CHAPTER 1

COMBINATIONS OF HIGH AND LOW FREQUENCY DATA IN MACROECONOMETRIC MODELS

L.R. Klein and E. Sojo, University of Pennsylvania

1. ECONOMETRIC PRACTICE

Econometricians who try to follow and project the overall economy as closely as possible ("Economy Watchers") Frequently base their main forecasts on macroeconometric models, supplemented by the frequent flow — almost daily – of indicative information. As soon as reports are prepared about some specific area of economic activity, they are released to the public. At the extreme, we have instantaneous market reports, originating with the start of the day at the international date line and moving with the sun to Tokyo, Hong Kong, Sidney/Melbourne, Singapore, Frankfurt/Paris, London, New York, Chicago, Los Angeles/San Francisco. These data cover both commodity and financial market reports.

On a monthly basis, there are national statistics for price indexes, wage rates, employment, unemployment, orders, shipments, inventories, construction, exports, imports, and many other indicators. Most macroeconometric models are based on social accounting statements that are prepared at either quarterly or annual intervals. It is possible to construct social accounts more frequently, but it has not been done on more than a fragmentary or sporadic basis. In practical sense, we should assume that quarterly models will be used for some time to come. They will use sample data from the quarterly social accounts, many of whose entries will be time aggregations of monthly or higher frequency data. The use of high frequency data will be for the construction of price deflators but also for the direct estimation of nominal (current value) entries in the accounts. The deflators of consumer spending will make use of

L.R.Klein ans Marquez (eds.), Economics in Theory and Practice:An Eclectic Approach,3-16.

monthly data on consumer prices; other deflators will use monthly prices from different indexes – wholesale, producers, commodity markets with the exception of government employee compensation which is obtained implicitly (nominal/real). Personal income series are actually produced monthly, and their entries in the national income accounts are obtained from monthly tax reports, social insurance records, company dividends when issued, and monthly public sector budget reports.

On the expenditure side of the social accounts, consumer outlays are computed from monthly (weekly) retail sales reports, inventory investment from monthly inventory reports and export/import outlays from monthly trade statistics, which need to be converted to a national account basis.

The preparation of the social accounts is a major statistical activity involving the combining of many diverse pieces of information. Many of these pieces are produced at high frequency intervals, but the total articulated report is quarterly, at the best, and for many countries only annually.

In using macroeconometric models for forecasting the national economy, practitioners are tied to the time schedule for preparing and releasing the social accounts. In the United States, the main lag is approximately three weeks. After the end of a calendar quarter, the first round estimates of the completed quarter are released on about the 20th of January, April, July and October. These lack profit statistics which become available one month later. Updating of data files and preparation of the forecast typically takes about 7 to 10 days, although it could be speeded up, if necessary. During the second and third months of a quarter, forecasts are examined and revised, partly an the basis of revisions in the quarterly social accounts on about the 20th of each month and partly on the basis of high frequency data which provide new information about short run economic movements. Official data – preparing agencies use high frequency data for the estimation of quarterly totals and must frequently round out a quarter's estimates by projecting missing months within the quarter.

The econometric forecaster using a quarterly model therefore starts the projections with input data for the previous quarter (lags) and solves for the current quarter. Some things are known about the current quarter at the time the solution is initiated. The solution typically has a future horizon of some twelve quarters, or thereabouts.

Should the econometric forecaster simply stand by the model result and ignore what is already known about the current quarter, or should the econometrician regard the high frequency data as added information, outside the sample? Since econometric samples are small and data are scarce, it seems sensible to make use of the added information. In practice, this is done by adjusting the constant term of relevant equations (or using a nonzero error term) in order to make the model outcome agree with the new information.

In a single equation this amounts to replacing an estimated equation value

$$y_t = f(x_t, \hat{\theta}) + 0 \tag{1}$$

where y_t = dependent variable, x_t = independent variable (input data), $\hat{\theta}$ = estimated parameter value

by

$$y_t = f(x_t, \hat{\theta}) + a_t \tag{2}$$

where a_t = adjustment value.

The original equation, before estimation, is

$$y_t = f(x_t, \theta) + e_t \tag{3}$$

In equation (1) the value of θ is replaced by its estimate $\hat{\theta}$, which is, in a sense, a representative value from a sampling distribution. In the same sense 0 is the mean of the distribution of e_t. If the function f is linear, this is obviously equivalent to changing the constant term of the equation by the amount a_t and leaving the zero mean of the distribution of e_t in place.

This widely practiced procedure has been criticised as being informal and subjective. A defense can be made for the practical procedure but an objective procedure that can be digitally replicated is available.

It has been observed that time series analysis of high frequency data provide good forecasts of economic magnitudes for the short run, say up to six months. Some analysts may claim validity for longer horizons, but experience with many series suggests good performance in short horizons. Such performance is not generally superior to model performance, but it is as good as model performance. It has also been observed and formally rationalised that combinations of forecasts, by different procedures, reduce the risk of forecast error, and may be preferred to forecasts from one method alone. In this spirit of inquiry we propose the estimation of current quarter forecasts by time series methods, using as much high frequency data as possible, and using these forecasts as benchmarks to which to adjust a quarterly model. The time series adjusted model is then extrapolated over a lenghier horizon. This approach has been independently pursued by econometricians at the Federal Reserve Board and the University of Michigan.[1] The present paper uses different high frequency models and different econometric models than those in other studies. This is, therefore, an independent investigation along lines similar to those followed elsewhere.

The idea of combining data of different frequencies in model analysis and forecasting is not new. In quite different contexts these combinations have been frequently made. In project LINK, where macroeconometric models from 79 countries or regions are simultaneously related to one another through trade flows, we have consistently combined quarterly models of the main industrial countries with annual models of the developing and centrally planned economies. Final results are reduced to the lowest common denominator, namely, annual data, but in getting the results, we first evaluate quarterly models, add or average quarterly values into annual values and then combine the time – aggregated data with the results from the annual models. This process needs to be iterated, and in re-entering the quarterly models on the i-th interation, we must disaggregate the final annual results obtained on the (i-1)st iteration.

In another application of mixed frequencies, we combine, at Wharton Econometrics, short term business cycle results from a quarterly model with medium term (10 years or more) results from an annual model. Careful forecasting of zigs and zags is applied to the quarterly models, and the quarterly results are aggregated (averaged). The annual model is then adjusted, as described above, to reproduce approximately the same yearly time path that the aggregated quarterly results follow. After the annual model traverses the short run business cycle span of three years, it takes a life of its own for the remainder of the simulation horizon.

2. THREE APPROACHES TO THE CURRENT QUARTER MODEL

To make use of monthly or weekly data in model construction, and to have a capability of generating current (and subsequent) quarter forecasts of major economic magnitudes, we approach the problem from three sides:

(i) the expenditure side

(ii) the income side

(iii) through unstructured empirical indicator relations.

For case (i), we consider the main entries on the expenditure side of national income and product accounts (NIPA) and also the main data sources used by the US Department of Commerce in constructing their accounts. We estimate time series and other interrelationships among high-frequency variables and then establish empirical "bridge" equations between the entries in the national income and product accounts and the high frequency data.

Where possible, the bridge equations are estimated from monthly data on both the indicators and the NIPA components. All NIPA components, however, are not reported monthly. In such cases, we must build quarterly bridge equations by aggregating the monthly indicators into quarters and correlating those with quarterly NIPA series. Our estimated equations are designed so as to be able to extrapolate the high frequency data series to fill out the current quarter and extend beyond for one or two more quarters.

In case (ii) we do the same thing for the main entries on the income side of the national income accounts that we do in case (i) for the expenditure side. In most of the prior work done at the Federal Reserve Board and elsewhere, the emphasis has been on the expenditure side. Of course, we have the problem of reconciling the two central aggregates produced by the methods of case (i) and case (ii) in reaching a single main aggregate to represent either gross national expenditure (product) or national income.

Case (iii) has no particular accounting structure. We simply put together all the "quick" information that we can assemble at high frequency early in the quarter. Much of it is intercorrelated. We then extract the leading principal components of the quarterly aggregates (averages) of these several indicators and regress major quarterly magnitudes on the quarterly aggregates (averages) of the prencipal components.

Let us first consider the detailed procedures followed in case (i), the construction of the current quarter estimates of the expenditure side. The Bureau of Economic Analysis of the United States Department of Commerce makes and releases its first estimates of a quarter's end. This is a preliminary estimate and is based on monthly or weekly data, much of which are incomplete for the entire quarter being estimated. Our procedure is to use the same high frequency data that the BEA use in making the preliminary forecast. In fact, a rigorous test of a model used for forecasting is to demand that the model be as "close" to the ultimate estimate as is the preliminary figure. After all, the estimate of the official national income statisticians represents a figure that is as close as we, who do not have access to all the details, could conceivably be able to come.

We first estimate nominal GNP and its components on the expenditure side. Some of these require a physical volume indicator and an associated price indicator. Next we estimate price deflators for categories of nominal (current value) expenditure items of the GNP. The real expenditure side estimates are obtained by deflating nominal expenditures by the estimated price deflator. The nominal GNP is built up from the following table, which lists the National Income and Product Account series on the left and the indicator series on the right hand side.

NIPA (quarterly)	**INDICATOR (monthly)**
Personal Consumer Expenditures	Retail Sales of:
Durable goods	
Autos and parts	Unit sales and CPI, autos
Furniture & household equipment	Furniture, home furnishings & equipment stores
Other durable goods	Durable goods less specific categories
Nondurable goods	
Food and beverages	Food stores
Clothing and shoes	Apparel and accessory stores
Gasoline and oil	Gasoline service stations
Other nondurable goods	Nondurable goods less specific categories
Services	Employment in services
Gross private domestic investment	
Nonresidential structures	Value of new nonresidential construction Put in place
Producers durable equipment	
Motor vehicles	Personal consumer expend., autos & Parts

Other equipment	Exports, machinery & transport. equip. Imports, machinery & transport. equip. Manufacturers' shipments of nondefense Capital goods
Residential structures	Housing starts and value of new Residential construction put in place
Nonfarm inventory change	
Manufacturing	Book value of manufacturing inventories
Retail trade	Book value of retail trade inventories
Merchandise wholesalers	Book value of merch. wholesale Inventories
Nonmerch. Wholesalers	ARIMA (autoregressive integrated moving average)
Other	ARIMA
Farm inventory change	ARIMA
Exports	Merchandise exports
Merchandise exports	Exports of goods
Service exports	Residual
Imports	Merchandise imports
Merchandise imports	Imports of goods
Service imports	Residual
Government purchases	
Federal	Net outlays, federal government
Federal structures	Federal government new construction put in place
Fed. Employee compens.	Federal government employment
Other federal	Residual
State & local employee compensation	State & local government employment
State & local structures	State & local govt. new construction put in place
Other state and local	ARIMA
<u>Deflators of the GNP</u>	<u>Consumer price indexes of:</u>
Durable goods	Durable goods
Autos and parts	New cars
Furniture & household equipment	Funiture & bedding and household appliances
Other durable goods	Residual
Nondurable goods	Nondurable goods
Food and beverages	Food and beverages
Clothing and shoes	Apparel
Gasoline and oil	Motor fuel, oil, and coolant
Other nondurables	Residual
Services	Services
Gross private domestic investment	
Nonresidential structures	PPI, intermed. materials for construction

Producers durable equipment	
Motor vehicles	ARIMA
Other equipment	Producer price of finished goods, capital equipment
Residential structures	PPI, intermed. materials for construction
Nonfarm inventory change	
Manufacturing	Implicit (nominal/real)
Retail trade	Implicit (nominal/real)
Merchandise wholesalers	Implicit (nominal/real)
Nonmerch. wholesalers	Implicit (nominal/real)
Other	Implicit (nominal/real)
Farm inventory change	Implicit (nominal/real)
Merchandise exports	Unit value of exports
Service exports	Merchandise exports
Merchandise imports	Unit value of imports
Service imports	Merchandise imports
Government purchases	
Federal employee compensation	Implicit (nominal/real)
Structures	Nonresidential structures
Other	Producer price index, capital equipment
State & local employee compens.	Implicit (nominal/real)
Structures	Nonresidential structures
Other	Producer price index, capital equipment

The data in this table are used in two ways. First, bridge equations are estimated. These are regressions of NIPA quarterly series on quarterly aggregates of the indicator series. Where there is no appropriate indicator available, the NIPA series is estimated directly from an ARMA process. For any NIPA—type variable, y_{it}, we estimate a quarterly ARMA process as

$$y_{it} + \sum_{j=1}^{p} \alpha_{ij} y_{i,t-j} = \sum_{j=0}^{q} \beta_{ij} e_{i,t-j} \qquad (4)$$

differencing y_{it} when the autocorrelations show signs of nonstationarity. The α_{ij} and β_{ij} are estimated by time series techniques, testing the random errors, $e_{i,t}$ for randomness; i.e.; we assume these random variables to be white noise errors.

In the second stage, after we have determined the regressions of the NIPA variables on indicators, where possible, we estimate monthly ARMA equations for each of the indicators. In applications of this system, we first extrapolate the indicators so that values for complete quarters (the current plus one or two ahead) can be estimated. We then put these estimated indicators on the right hand side of the bridge equations to obtain values for the NIPA variables in the current and projected quarters.

A typical equation is

$$\text{CENGD} = 0.0403 + 9.470\ \text{ICENGD} \qquad (5)$$

$$(1.01) \quad (23.84)$$

$$R^2 = 0.71, \quad \rho_1 = -0.13$$

CENGD = Consumer expenditures on gasoline and oil in current prices and in first differences

ICENGD = Retail sales of gasoline service stations in current prices and in first differences

The correlograms of residual error is flat, near zero, and the first order serial correlation is –0.13.

The indicators are generally highly correlated with the corresponding NIPA variable. All the equations of the system are listed in an annex.

The GNP and various sub--categories are obtained from the relevant sums (or differences) of the components. The components in constant prices are estimated by dividing the estimated nominal series, listed above, by the estimated price deflators. The deflators themselves are estimated in the same way that the current value series in the GNP are estimated.

The design of calculations on the income side proceeds in an analogous way. We begin with a table of relationships between NIPA and indicator elements.

NIPA (quarterly)	**INDICATORA (monthly)**
Gross national product	From identity
Less: capital consumption allowances	Fixed investment (see expenditure side)
Equals: net nation. product	From identity
Less: business transfer payments	Residual
Less: statistical discrepancy	Residual
Plus: subsidies less current surplus	Net outlays, federal government (see expenditure side)
Equals: national income	From identity
Less: corporate profits & adjustments	Industrial production Retail sales
Less: contributions for social insurance	
Personal	Wage & salary disbursements
Employer	Wage & salary disbursements or personal contributions for social insurance
Less: wage accruals less disbursement	ARIMA
Plus: transfer payments	
Government	Total transfer from monthly personal income
Business	Residual
Plus: interest paid by government & consumers	Outstanding consumer installment credit federal governmt. debt, interest rate, prime comm. paper, 6 months
Net interest paid by consumers	Outstanding consumer installment, credit, interest rate, prime comm. paper,6 months
Net interest paid (fed)	Federal governmt. debt, interest rate, prime comm. paper,6 month
Less: interest paid to foreign	Resicual
Net interest paid (state) Plus: dividends	Personal dividends from monthly pesonal income
Equals: personal income	From identity
Personal income	From identity
Wage & salary disbursements	
Commodity producing industries	
Manufacturing	Employment manufacturing avg. weekly,hours, manufacturing average

	hourly,earnings
Other	Employment, other average weekly hours, manufacturing average hourly earnings
Distributive industries	Employment, wholesale & retail trade avg. weekly hrs, wholesale & retail trade avg. hourly earnings
Services	Employment, services average weekly hours, services average hourly earnings, services
Government	ARIMA
Other labour income	Employment
Proprietor's income	
Farm	Prices paid by farmers Prices received by farmers
Business & professional	Industrial production retail sales
Rental income of persons	ARIMA
Personal dividends	Dividend to price ratio, stock prices
Personal interest income	
Interest paid by consumers	Outstanding consumer installment credit interest rate, prime comm. paper, 6 months
Other	Interest rate, prime comm. paper, 6 months
Transfer payments to persons	Total unemployment
Less: personal tax & nontax paymts.	Wage and salary disbursements
Equals: disposable personal income	From identity

The income side is naturally estimated in current prices. The values could be deflated by some combination of price indexes that are simultaneously projected from the expenditure side, and the deflated values could be taken as representative of real GNP. It is, however, less speculative to confine the estimates from the income side to current values and to compare such estimates for reconciliation purposes with current value estimates obtained from the expenditure side.

In the third case, we aim to use as much high frequency data as are available in the current quarter to estimate just three magnitudes – real GNP, nominal GNP, and the GNP deflator. Since these are related by the identity

PDGNP * GNP = GNP ($)

only two series should be computed directly from indicator data; the third should be obtained from the above definitional equation. Early readings on a number of variables are collected monthly. These are meant to be revealing about the overall economy.

Enough values are estimated for the current quarter so that we have three monthly figures (observed and projected) for each indicator variable. Leading principal components are computed from the observed set of quarterly aggregates of monthly variables. The three aggregates (PDGNP, GNP, and GNP($)) are regressed on the leading principal components. Projected values of the indicator variables for the current quarter and also for one quarter ahead are substituted into the regression on principal components in order to obtain current and future quarter forecasts of the desired aggregate.

The source data used to form indicators of PDGNP, GNP and GNP($) are listed below.

INDICATORS (MONTHLY)
PDGNP, Price deflator of GNP
Consumer price index
Producer price index, finished goods
Producer price index, intermediate materials
Average hourly earnings
Average weekly hours
Unit value index of imports
Prices received by farmers

GNP($), Nominal gross national product
Value of shipments, manufacturing
Value of new orders, manufacturing
Value of unfilled orders, manufacturing
Nominal personal income
Money stock (M1)
Retail sales
Interest rate on 6 month CDs
Index of net business formation

GNP, Real gross national product
Value of shipments deflated by PPI intermediate materials
Value of new orders deflated by PPI intermediate materials
Value of unfilled orders deflated by PPI intermediate materials
Nominal personal income deflated by CPI
Money stock (M1) deflated by CPI
Retail sales deflated by CPI
Real interest rate (6 month CD rate less CPI inflation rate)
Industrial production index
Employment
Average weekly hours

Each indicator variable is projected ahead by an ARMA process that has been fitted to historical data on the monthly indicator values.

Each indicator is first fitted to the semilog trend formula

lnI = a + bT

I = quarterly aggregates of monthly indicator
T = chronological time
a,b = estimated regression coefficients

Detrended values are computed from

I/exp(a+bT)

Principal components are extracted from the detrended variables after standardisation. Call the components $(PC)_1$, $(PC)_2$, etc.

We fit the empirical regressions

$$PDGNP^* = c_0 + \sum_i c_i (PC)_i$$

$$GNP(\$)^* = d_0 + \sum_i d_i (PC)_i$$

$$GNP^* = e_0 + \sum_i e_i (PC)_i$$

The asterisk denotes that detrended values are used for dependent variables in the three regressions.

The residuals are examined for serial correlation and well—known autoregressive corrections are used in order to isolated white noise errors from the principal component regressions.

Detrended estimates of the main variables are obtained and then transformed back to original units with trend. The mean absolute percentage errors of estimate over the sample period are

real GNP	0.28
nominal GNP	0.39
price deflator	0.21

3. APPLICATIONS

The use of the three different current quarter models (expenditure side, income side, and empirical indicator model) is self evident. As data become available during the first or second month of any quarter, estimates of the entire quarter and the next future quarter are made. This could be done at any time, but typically during the early days of the second month. The three models could be averaged. The expenditure and income side models could be brought close together by minimising the one reconciling item, the statistical discrepancy.

These estimates can be used for the early quarters of a forecast, but they can also be used for the purpose of adjusting, or calibrating, a model so that it starts from realistic values. In the process of making adjustments to the constant or error terms of equations in a model, it has been customary to bring the model into agreement with observations for period T and T-1, the latest observed periods just before the forecast projection period, T+1,T+2,T+3, …

The values of a_t in equation (2) above are the adjustment values. They can be chosen so that y_t, the variables that are to be forecasted, agree with the values of the current quarter model. In other words, choose a_t such that

$$y_t^* = f(x_t, \hat{\theta}) + a_t$$

where

y_t^* = forecast of y_t from the current quarter model

$f(x_t, \hat{\theta}) =$ value of y_t computed from the estimated structural equation solved in a deterministic mode.

Since y_t^* are computed from objective equations that can be replicated, this is an objective adjustment procedure.

Some forecasters have a tendency to adjust many equations of a model – substantially more than one—half.

Our view is that adjustment should be kept to a minimum. This means that only the equations that are mainly responsible for leading variables (like GNP) should be adjusted. The results should be scrutinised to be sure that no variable is terribly far out of line.

In principle, it would be possible to construct a loss function so that a selected number of adjustment values were chosen in order that the loss value is minimised.

$$F(y_t', y_{t-1}', \ldots\ldots, x_t', x_{t-1}', \ldots\ldots, \hat{\theta}') = a_t$$

subject to:

$$F(y_t', y_{t-1}', \ldots\ldots, x_t', x_{t-1}', \ldots\ldots, \hat{\theta}') = a_t$$

where F is a column vector of functions, y_t is a column vector of endogenous variables (equal in number to the count of elements of F), and x_t is a column vector of exogenous variables. x_t is a column of parameter estimates, and a_t is a column vector of adjustment values. The weights w_i in L, show the relative importance of the various elements of y that are evaluated in the current quarter model. These weights should reflect the variables for which the current quarter model shows greater possibilities of improving the econometric model and may vary with the monthly observations we have about the current quarter. The problem is to search for values of a_t, such that the solution of the entire system brings y_t as close as possible, in the square, to y_t^*. Mny of the elements of a_t could, of course, be zero. In some models, it will be possible to bring L to zero. This is where it is possible to bring y_t to their targets y_t^* without violating any equations of the system. In general, however, if all the values of y_t are fixed exactly at y_t^*, it may not be possible to find a consistent solution of the system. In this case, we seek a least-squares solution for deviation of y_t. In any event, if the number of nonzero elements of a_t is to kept quite small, there will be fewer nonzero elements of a_t than there are target values y_t^* of y_t. In such cases, the objective will be to bring y_t near to y_t^*, in the sense of minimum squared deviation.

It should also be possible to chose values of selected elements of x_t in order to minimise L, leaving all $a_t = 0$. In most cases, we shall have a priori information about x_t for the short run – up to six months – and will not, for the present problems, want to search for optimal values of x_t.

NOTES

1. Carol Corrado and Mark Greene (1988), "Reducing Uncertainty in Short—Term Projections: Linkage of Monthly and Quarterly Models", journal of Forecasting 7, pp.77—102. Mark Greene, E. Philip Howrey, and Saul H. Hymans, "The Use of Outside Information in Econometric Forecasting", in D.A. Belsley and E. Kuh (eds.) (1986), Model Reliability, Cambridge: MIT. The estimation of current quarter models from high frequency data was initiated many years ago by Otto Eckstein at Data Resourses, Inc.