6 The LIMA forecasting model of the Institute for Advanced Studies Vienna

The LIMA model has grown out of the LINK project that attempts to join worldwide economic forecasting models into a common framework. Because many of the variables are only available at an annual frequency, the LIMA model also operates at this annual frequency. This can be troublesome for short-run prediction, as inofficial provisional data on main accounts aggregates come in on a quarterly basis. Therefore, LIMA is rarely run in its original form with zero residuals, and 'add factors' play a key role.

The LIMA model is a traditional macroeconometric prediction model with an emphasis on the economy's demand side. Thus, the model may be called a 'Keynesian' model. It has 75 equations, which implies 75 endogenous variables. As in most macroeconometric models, most equations are mere identities. Only 21 equations are 'behavioral' and contain estimated coefficients. With 75 endogenous variables and 21 structural equations, the LIMA model is a comparatively small macroeconometric model.

LIMA's model structure is updated frequently. Some equations may be replaced by better ones, while others are being eliminated in a search to simplify the overall structure and again others are being added in order to satisfy the needs for a more refined modeling of a certain sector of the economy. In the recent decade, the needs for disaggregated modeling in many sectors has decreased on average. Accordingly, the 'true dimension' of the model has shrunk from a maximum of more than 30 to 21. The 'full dimension' was even larger, due to a very elaborate bookkeeping model for the Austrian public sector. Notwithstanding all modifications, the basic LIMA model structure still resembles its predecessors from the late 1970's.

Parameter estimates are updated once a year, when the official provisional data for the previous year become available, which is usually in September. For example, in September 2002 all equations were re-estimated using data from 1976 to 2001. For many equations, estimation intervals ended in 2000 if provisional data were earmarked as likely to be revised further. 1976 is the earliest year, for which national accounts data are available that correspond to the ESA standard. With some temporary exceptions, all equations are estimated by OLS. Indications of mis-specification due to autocorrelation are adjusted by dynamic modeling rather than by GLS–type corrections. Thus, most behavioral equations are dynamic.

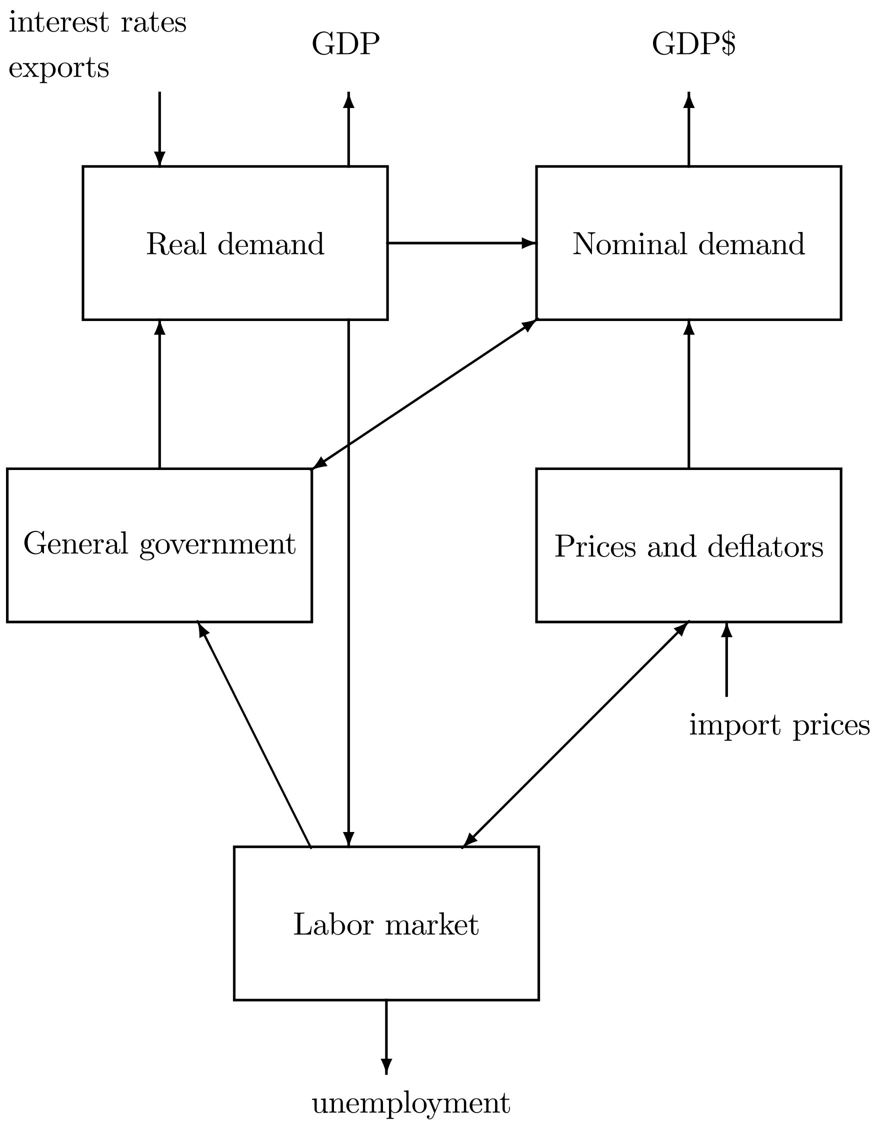

Figure 1: Structure of the forecasting model LIMA.

The model's center piece is the *domestic demand sector.* Demand aggregates are modeled in real terms, i.e. at constant prices, and sum up to real gross domestic product (GDP). Additional equations are used to determine *prices and deflators.* By multiplying those deflators into the real aggregates, nominal variables and eventually nominal gross domestic product (GDP$) are calculated.

This adding-up to obtain GDP requires export and import variables. The treatment of exports and imports is asymmetric. *Imports* are fully endogenous and respond to demand categories, such as consumer durables and equipment investment. By contrast, exports are mainly exogenous. Older LIMA versions considered modeling exports as depending on world demand but, unfortunately, data on world demand become available with a considerable time lag only, which excludes its usage for the practical purpose of forecasting. For export and import prices, the approach is reversed. Import prices are exogenous, as it is assumed that Austrians have to accept the world market's price level, while export prices are endogenous.

Another component of GDP is public consumption. In the current version, public consumption is exogenous. In earlier versions, nominal public consumption was modeled as resulting from the sum of spending categories of general government. This practice was abandoned, as most government spending categories are exogenous and as the resulting price deflator of public consumption was often implausible or caused instabilities in model solution. In contrast to spending, several components of government revenues are modeled as endogenous variables, such as direct taxes or contributions to social security. From this *government sector*, balancing items such as the budget deficit can also be calculated.

The real and government sectors also interact with the *labor market sector*, which yields variables such as employment, the labor force, and wages. Other variables, such as the working-age population, are exogenous.

The LIMA model does not include a *financial sector.* Financial variables that are influential for the goods market, such as exchange rates and interest rates, are supplied by specialists on the financial sector who use separate models.

6.1 A typical demand equation: consumer services

Consumer demand consists of three categories: consumer durables, consumption of other goods, and consumer services. Almost 50% of household expenditures are spent on services. The share of services in household consumption appears to be increasing in the longer run. Before 1980, it used to be below 45%.

As a general rule, demand equations use logarithmic growth rates as dependent variables. Logarithmic growth rates are fairly constant in the longer run, hence they come closer to fulfilling the assumption of stationarity than, for example, first differences. On the other hand, percentage growth rates are far less convenient to handle from an econometric model builder's viewpoint.

In all consumption equations, the principal explanatory variable is the growth rate of household disposable income yd. The real variable yd is obtained from deflating nominal household income by the consumption deflator. Therefore, the price index of total consumption deflates income, while a special price index for consumer services deflates the dependent variable. It is tempting to explain the demand for services by a relative price, reflecting the idea that services and goods are partial substitutes. However, such attempts fail to yield significant explanation.

Another potential source of explanation comes from error-correction relationships. While economic theory and plausibility dictate that the long-run elasticity of consumption with respect to income should be one, this is not so for consumer subaggregates. For example, a cointegrating regression of log consumer services on log income

$$cs_t = b_0 + b_1 yd_t + u_t$$

yields $\hat{b}_1 = 1.1984$, slightly in excess of unity, while the comparable value for non-durables is below one. In theory, unit elasticity for total consumption should be imposed on the model. This is technically difficult, however, due to the implied non-linear restriction structures. Therefore, this important long-run restriction is ignored in estimation. The cointegrating regression is estimated by least squares, and the resulting error-correction variable $cs - \hat{b}_1 yd$ is then used as an additional regressor.

The estimation results are acceptable. All regressors are significant, and the (here, not very reliable) Durbin-Watson statistic does not indicate any serious specification error. The R^2 is lower than in other consumption equations, which would be an incentive to search for possible further explanatory

Table 1: Behavioral equation for consumer services. Estimation time range is 1967–2002. Dependent variable is $\log(cs_t/cs_{t-1})$.

regressor	coefficient	t–value
constant	-0.382	-2.858
$\log(cs_{t-1}) - 1.198 * \log(yd_{t-1})$	-0.227	-3.023
$\log(yd_t/yd_{t-1})$	0.330	3.372

$R^2 = 0.428$, DW=1.638

variables. Unfortunately, neither interest rates at any lags nor lags of the dependent variable yield a significant explanatory contribution.

It is less comforting that the last sample observation in 2002 yields a large negative residual. In other words, consumer services fell much below the predicted value. This is quite inconvenient for prediction. The low value of 2002 tends to 'boost' the forecast for 2003 because of the error-correction influence, maybe beyond plausible limits. Indeed, the current IHS forecast uses a negative residual add factor for 2003. Of course, at the time of writing a provisional value for 2003 consumption is known already, thus there is an *ex post* justification for this negative add factor.

6.2 Overruling statistical evidence: equipment investment

Besides consumption, investment or 'gross fixed capital formation' is another important component of aggregate demand. While the ESA system disaggregates investment into a larger number of subcomponents, LIMA only considers equipment investment, which includes machinery and transportation equipment, and construction investment, which includes business as well as residential construction. Equipment investment is the slightly smaller part but its equation is more important than the construction investment counterpart, as construction often relies much on public funding and policy.

While the basic idea for consumption modeling is dynamic error correction, investment demand relies on factor demand equation that are typically derived from assumed Cobb-Douglas or CES production functions. In these concepts, a primary determinant of investment is current output growth, which is interpreted as indicating the short-run tendency in demand that should be satisfied by production, for which in turn investment is necessary.

Another influential regressor is an adjustment term that may be read as a cointegrating influence. Using the logged share of equipment investment in total output as a regressor assumes that the share of equipment investment in total output is fairly constant in the longer run. This is not necessarily true and is not really backed by theory. Economic theory yields a constant share of *total* investment in output only. A cointegrating regression yields an elasticity of 1.33, which appears to be too high for forecasting purposes. Therefore, the assumed unit elasticity is imposed instead of the estimated number.

Economic theory suggests a negative influence from real interest rates on investment demand. Unfortunately, such an influence is not backed by empirical evidence. A lengthy search among various constructions for real interest rates resulted in a long-term ten-year rate that is deflated by investment prices. Even this 'optimum regressor' fails to achieve convincing significance. While statistical evidence suggests removing the link from interest rates to investment, following this suggestion could be inconvenient, particularly for conditional forecasting and scenarios.

Table 2: Behavioral equation for equipment investment. Estimation time range is 1980–2001. Dependent variable is $\log(ife_t/ife_{t-1})$.

regressor	coefficient	t–value
constant	-0.710	-3.150
d8283	-0.110	-4.174
$\log(ife_{t-1}/y_{t-1})$	-0.311	3.291
$\log(y_t/y_{t-1})$	2.597	4.459
real interest	-0.010	-1.464

$R^2 = 0.712$, DW=2.068

A sizeable aberration requires the usage of a dummy variable for two years in the early 1980's. The introduction of such dummy variables should be restricted to occasions where they are absolutely necessary. The shorter time range of available data for the interest rate results in a shortened estimation interval for this equation. In summary, degrees of freedom are few in this case.

6.3 An example for a deflator equation: investment prices

For each demand aggregate, two behavioral equations must be specified: an equation for real demand and an equation for the price deflator. In the case of equipment investment, the corresponding price deflator is named $pife$, for '$p$rices of $i$nvestment $f$ixed $e$quipment'. A large part of equipment investment demand is satisfied by imported goods, therefore the price deflator should be influenced directly by import prices. Another explanatory variable is ulc, unit labor costs, which stems from the labor market sector of the LIMA model. Substantial autocorrelation in the deflator also requires the insertion of lags. Thus, the $pife$ equation is a severely dynamic regression equation. As a general rule, dynamic equations support the stability of the model, while static equations may result in unstable behavior. Therefore, in spite of some indication of remaining residual correlation, the $pife$ equation is satisfactory.

Table 3: Behavioral equation for the deflator of equipment investment. Estimation time range is 1978–2001. Dependent variable is $\log(pife_t/pife_{t-1})$.

regressor	coefficient	t–value
$\log(pife_{t-1}/pife_{t-2})$	0.318	1.810
$\log(ulc_{t-1}/ulc_{t-2})$	0.243	2.379
$\log(pmg_t/pmg_{t-1})$	0.156	2.157
$\log(pmg_{t-1}/pmg_{t-2})$	0.104	1.312

$R^2 = 0.660$, DW=2.471

The large value of the Durbin-Watson statistic should be seen against the backdrop of its tendency to be biased toward the ideal value of two in dynamic regression. Thus, the evidence on negative autocorrelation is stronger than would otherwise be indicated by a value of 2.47. Such negative residual correlation may point to an over-fit caused by too many regressors or it may be caused by the omission of the constant. Here, the effect is rooted in the early years and it is mainly due to the effects of the dummy variable. It is to be noted that this equation, in line with most price equations, does not have a constant term. This implies that individual demand aggregates do not have an inflationary core of their own but that they just pick up price developments of their inputs.

6.4 Employment: no more Phillips curve

The employment equation ranges among the most frequently modified model equations. Earlier versions often included inflation among the regressors, while the present specification relies on error correction and on relative prices. The main determinant of employment, however, is real output growth. The coefficient on real output growth shows the effects that are otherwise known as Okun's law.

Table 4: Behavioral equation for employment excluding self-employment. Estimation time range is 1981–2001. Dependent variable is $\log(le_t/le_{t-1})$.

regressor	coefficient	t–value
constant	0.447	3.775
d83	-0.020	-3.662
$\log(gdp_t/gdp_{t-1})$	0.375	4.173
$\log(le_{t-1}/gdp_{t-1})$	0.301	3.804
$\log(ywgle_{t-1}/pgdp_{t-1})$	-0.348	-3.767

$R^2 = 0.763$, DW=1.833

All regressors are significant and have the expected signs. Unfortunately, the inclusion of a dummy variable was necessary. Fortunately, it is located in the earlier years and may have only small effects on forecasting.

The short-run Okun-type coefficient has the plausible value of around 0.4. Note that it is not exactly the same as in Okun's law, due to some non-linear transformations and due to the omission of the labor-supply effects that are also captured in the original Okun coefficient. Error correction has a sizeable impact, which implies that the long-run unit elasticity shows its effects after fey years already. In other words, a sudden recession has only small effects on employment, while the full negative effects are felt if the recession does not end soon.

The negative effects of real wages, i.e. the relative price of the production factor labor, are also quite strong. The variable *ywgle* is the *per capita* gross wage. Technically, it counteracts the tendency of employment to grow proportional to output, which would imply an absence of technological progress. However, the long-run growth of real wage puts a brake on unlimited employment expansion. Thus, the employment equation is a stabilizing component in the LIMA model, even though its structure will certainly have to be reconsidered from time to time.

6.5 Theory and practice and the LIMA model

For many years, the LIMA model has proved a valuable backbone of the official IHS economic forecast. According to most comparative evaluations, the IHS economic forecast and that of its main competitor, the WIFO Institute for Economic Research, are of a comparable quality. Usually, forecasts are published for the current and for the following year only. Once a year, a medium-term projection is also presented to the public.

The following points can be identified where the LIMA forecast does not correspond to textbook forecasting:

1. There is no sharp boundary between a sampling interval and a prediction interval. Most macroeconomic variables—exceptions are exchange rates and unemployment data—spend years in an intermediate stage, where they are known with an increasing degree of precision.

2. Often, predicting the final data, i.e. those that mark the endpoint of all revisions by statistical agencies, are *not* targeted. For example, a forecast for 2000 may become uninteresting in 2003, even when it perfectly coincides with the final value.

3. The basis for prediction does not coincide with the estimation interval. For example, a forecast for 2004 is based on provisional data for all lagged variables from until 2003, while model parameters have not been updated from values beyond 2002 or even 2001.

4. Add factors play a key role. Incoming information is reflected in sizeable adjustment of residuals. Usually, zero-residuals forecasts are far off the mark.

5. Exogenous variables for the prediction interval are updated on an *ad hoc* basis. Some of these, however, correspond to information provided by other institutions—for example, government spending—or by researchers who use separate forecasting models.

6. All estimation is conducted by OLS, even when it is known that this procedure yields inconsistent estimates.

These points should not be interpreted as a critique of the current practice. Rather, one may assume that the discrepancies between the textbook prediction approach and current practice follow a longer-run experience of forecasters on how to do forecasting efficiently. It may be interesting to extend the textbook framework, taking the practitioners' approach into account. This direction of research is still in an early stage.

References

[1] Anderson, T.W. (1951) 'Estimating linear restrictions on regression coefficients for multivariate normal distributions'. *Journal of the American Statistical Association* **85**, 813–823.

[2] Box, G.E.P., and G.M. Jenkins (1970) *Time Series Analysis, Forecasting, and Control.* Holden-Day.

[3] Brockwell, P.J., and R.A. Davis (1991) *Time Series: Theory and Methods.* 2nd edition, Springer-Verlag.

[4] Cartwright, N. (1995) 'Probabilities and experiments'. *Journal of Econometrics* **67**, 47–59.

[5] Chatfield, C. (2001) *Time-series Forecasting.* Chapman & Hall.

[6] Christoffersen, P.F., and Diebold, F.X. (1998) 'Cointegration and long-horizon forecasting', *Journal of Business & Economics Statistics* **16**, 450–458.

[7] Clements, M., and Hendry, D.F. (1998) *Forecasting economic time series.* Cambridge University Press.

[8] Dickey, D.A., and Fuller, W.A. (1979) 'Distribution of the estimators for autoregressive time series with a unit root' *Journal of the American Statistical Association* **74**, 427–431.

[9] Engle, R.F. (1992) 'Autoregressive conditional heteroskedasticity with estimates of variance of United Kingdom inflation' *Econometrica* **50**, 987–1007.

[10] Engle, R.F., Hendry, D.F., and Richard, J.-F. (1983) 'Exogeneity'. *Econometrica* **51**, 277–304.

[11] Engle, R.F., and Yoo, B.S. (1987) 'Forecasting and Testing in Co-integrated Systems', *Journal of Econometrics* **35**, 143–159.

[12] Franses, P.H. (1996) *Periodicity and Stochastic Trends in Economic Time Series*. Oxford University Press.

[13] Franses, P.H., and VanDijk, D. (2000) *Non-linear time series models in empirical finance*. Cambridge University Press.

[14] Gardner, E.S. Jr. (1985) 'Exponential Smoothing: The State of the Art' *Journal of Forecasting* **4**, 1–28.

[15] Gardner, E.S.Jr., and McKenzie, E. (1985) 'Forecasting trends in time series' *Management Science* **31**, 1237–1246.

[16] Granger, C.W.J. (1969) 'Investigating Causal Relations by Econometric Models and Cross-Spectral Methods' *Econometrica* **37**, 424–438.

[17] Granger, C.W.J. (1989) *Forecasting in Business and Economics*. Academic Press.

[18] Harvey, A.C. (1989) *Forecasting, Structural Time Series, and the Kalman Filter*. Cambridge University Press.

[19] Hylleberg, S., Engle, R.F., Granger, C.W.J., and Yoo, B.S. (1990) 'Seasonal integration and cointegration' *Journal of Econometrics* **44**, 215–238.

[20] Johansen, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

[21] Pindyck, R.S. and Rubinfeld, D.L. (1991). *Econometric Models and Economic Forecasts*. 3rd edition, McGraw-Hill.

[22] Sims, C.A. (1980) 'Macroeconomics and reality'. *Econometrics* **48**, 1–48.

[23] Taylor, S. (1986) *Modelling Financial Time Series*. John Wiley & Sons.

[24] Tong, H. (1990) *Non-linear Time Series: A Dynamical Systems Approach*. Oxford University Press.