Alternatives to the Autoregressive Moving Average Structure in State Adjustment Dynamic Demand Modeling

by

Steven C. Fischer

Inter-American Development Bank (retired)

10500 Rockville Pike #602
Rockville, MD 20852-3340
Phone: 301-564-9508/Fax: 301-564-1261/Email: fischer.s.c@gmail.com

Abstract

Conceptual exploration of several simple potential estimation approaches (with well-behaved disturbances) applicable under conditions of limited data availability to the state adjustment dynamic model (interpreted as a type of rational distributed lag dynamic model with the rational distributed lag operating on the lagged dependent variable).
The Houthakker and Taylor (1966, 1970) state adjustment demand modeling approach entails an autoregressive moving average (ARMA) structure that, in its simplest, first-order form [with the basic structural equation, featuring the simplest form of rational distributed lag on the lagged dependent variable, expressed in linear-in-logs rather than linear form such that short- and long-term explanatory variable elasticities equal the derivatives developed in Houthakker and Taylor -- see Breunig (2008)], yields the following model when applied to the annual time-series variables [\(\ln(\text{iec})\), \(\lnP\), \(\lnvai\) (logarithm of industrial electricity consumption, industrial electricity rate, and industrial value added, respectively), \(\protecto\) (mean percentage rate of import tariff protection), and time trend \(t\)] utilized in the industrial electricity demand assessment of Fischer (2002):

\[
\ln(\text{iec}_t) = B_1*(1 - B_7) + B_3*(\lnP_t - B_7*\lnP_{t-1}) + B_4*(\lnvai_t - B_7*\lnvai_{t-1}) + B_5*(\protecto_t - B_7*\protecto_{t-1}) + B_6*(1 - B_7)*t + (B_2 + B_7)*\ln(\text{iec}_{t-1}) + \epsilon_t - B_7*\epsilon_{t-1}
\]

or \(\ln(\text{iec}_t) = A_1 + A_3*\lnP_t + A_5*\lnP_{t-1} + A_4*\lnvai_t + A_4*\lnvai_{t-1} + A_5*\protecto_t + A_5*\protecto_{t-1} + A_6*t + A_2*\ln(\text{iec}_{t-1}) + \nu_t\)

where the disturbance \(\nu_t\) is a first-order moving average MA(1) with parameter \(B_7\); \(B_7\) is the first-order state adjustment parameter (i.e., the parameter of the denominator polynomial of one degree in the lag operator of the Jorgenson (1966) rational distributed lag on the lagged dependent variable in the basic structural equation); \(1-B_7\) is the first-order state ‘depreciation’ rate; the coefficient on the lagged dependent variable equals \(B_7\) plus \(B_2\), the parameter on the lagged state variable in the structural equation (equal to the numerator polynomial of zero degree of the rational distributed lag); in the case of the inertia and habit formation in electricity consumption, \(B_2\) is assumed to be positive; for \(B_7 = 0\), the model reduces to a partial adjustment model with a well-behaved disturbance.

To estimate the state adjustment model directly applying the maximum likelihood technique would be problematic, despite the sophistication of available econometric software, because of the limitations on data availability documented in Fischer (2002) (annual time series over 1966-1997) in the context of moving average errors and the associated complexity of the log-likelihood function. Thus, the problem of estimating this model focuses on the moving average disturbance term, which is derived from the underlying Houthakker-Taylor state adjustment structural framework. When this problem is systematically addressed at a preliminary conceptual level, several simple theoretical alternatives to the direct maximum likelihood estimation technique emerge, and are worth exploring empirically in the future.

Because of the lagged dependent variable, consistent estimation of this model by nonlinear or ordinary least squares (LS) (with independent coefficients estimated for the
lagged explanatory variables as in the second form of the model, above) is only possible if, in the disturbance \(v_t, \epsilon_t = B_7^* \epsilon_{t-1} + u_t \) where \(u_t\) is normally and independently distributed (well-behaved), that is, \(\epsilon_t\) follows an autoregressive (AR(1)) scheme whose parameter takes on the particular value \(B_7\), which is difficult to justify.\(^1\) Even if this assumption were valid, the estimation/data experience of Fischer (2002) is that, because of the lagged dependent variable \(\ln(\text{iec}_{t-1})\), LS (and nonlinear LS) estimation of models of this sort is affected by finite sample bias.\(^2\) Also, the instrumental variables method is an option, but experimentation with similar even simpler models in Fischer (2002) yields no success in the application of appropriate instruments for the lagged dependent variable, which may well depend on a considerable expansion of the data. Thus, we are still looking for a simple way to deal with the moving average disturbance term.

On the basis of a set of conceptual cases, Carter and Zellner (2003) suggest that it is always possible to re-express an autoregressive moving average model with a moving average disturbance in an alternative, distributed lag form that has comparable theoretical plausibility, a similar and/or related lag structure, and a well-behaved disturbance. Starting here, dynamic demand specifications are explored in the sections and annexes that follow that are (i) direct transformations of the original ARMA model and (ii) rough equivalents or approximations (Annex C) to these models. Another simple method for estimating the state adjustment model is set out in (iii) based on the hypothesis that the generalized moving average disturbance term is subject to first-order autocorrelation.

(i) **Klein/Zellner-Geisel Maximum Likelihood Method via Nonlinear Least Squares**

The first idea is to employ the method developed by Klein (1958), as operationalized by Zellner and Geisel (1970), entailing maximum likelihood estimation via nonlinear least squares, for the direct transformation of the above ARMA specification by dealing with the first-order moving average disturbance structure in the first equation above as follows:

\[
\begin{align*}
\text{E} & \ln(\text{iec}_t) = B_1^*(1 - B_7) + B_3^*(\ln P_t - B_7^*\ln P_{t-1}) + B_4^*(\ln \text{vai}_{t} - B_7^*\ln \text{vai}_{t-1}) \\
+ B_5^*(\text{protecto}_t - B_7^*\text{protecto}_{t-1}) + B_6^*(1 - B_7)^*t + B_7^*\ln(\text{iec}_{t-1}) + B_7^*\text{E} \ln(\text{iec}_{t-1})
\end{align*}
\]

\(^1\) If the parameter, say \(B\), of this model’s first-order moving average disturbance term equaled the coefficient on the lagged dependent variable, as in the case of adaptive expectations models, then a recursive transformation applied to this model and disturbance term would produce a well-behaved disturbance term \(\epsilon_t\) (to be exact, \(\epsilon_t B^*c_0\)), enabling nonlinear least squares maximum likelihood estimation with or without the approximation that \(\ln(\text{iec}_{t0})\) in the truncation remainder is fixed, equal to its expected value, i.e., that \(c_0\) is zero. Though mathematically convenient, this is not a real case because the moving average disturbance term with parameter \(B_7\) arises precisely because of the presence of the state adjustment effect and a distinct parameter \(B_2\) on the underlying lagged state variable, so that the parameter of the moving average disturbance term cannot possibly equal the coefficient on the lagged dependent variable.

\(^2\) The maximum likelihood method is applicable if \(v_t\) is normally distributed with zero mean and constant variance. The very special assumption that \(v_t = u_t\) suggests the estimation of a recursive transformation of the above model that eliminates the lagged dependent variable \(\ln(\text{iec}_{t-1})\) and creates a truncation remainder of the partial-adjustment-model type, but induces a disturbance term weighted by a finite lag scheme on \(A_{27}\). The autocorrelation associated with this disturbance and its impact on the significance of nonlinear least squares maximum likelihood estimation results can be ignored if \(A_{27}\) turns out to be a relatively small proportion. However, since \(A_{27}\) represents the sum of \(B_2\) and \(B_7\), this may very well not be the case.
leading to (after recursive transformation)

\[
\ln(iec_t) = B_1 + B_3*(\ln P_t - B_7*\ln P_{t-1}) + B_7*(\ln P_{t-1} - B_7*\ln P_{t-2}) + B_7^2*(\ln P_{t-2} - B_7*\ln P_{t-3}) + \ldots + B_7^{t-1}*(\ln P_1 - B_7*\ln P_0) + B_4*[(\ln inv_{1t} - B_7*\ln inv_{1-1}) + B_7*(\ln inv_{1-1} - B_7*\ln inv_{1-2}) + B_7^2*(\ln inv_{1-2} - B_7*\ln inv_{1-3}) + \ldots + B_7^{t-1}*(\ln inv_{1-0})]
\]

\[
+ B_5^2*[(\ln P_0 - B_7*\ln P_{-1}) + B_7^2*(\ln P_{-1} - B_7*\ln P_{-2}) + B_7^3*(\ln P_{-2} - B_7*\ln P_{-3}) + \ldots + B_7^{t-1}*(\ln P_{-0} - B_7*\ln P_{-1})]
\]

\[+ B_6^t + B_2^*\ln(iec_{t-1}) + B_7^2*\ln(iec_{t-3}) + \ldots + B_7^{3}*(\ln(iec_{t-3}) + \ldots + B_7^{t-1}*(\ln(iec_{0}))
\]

\[+ [\ln(iec_{0}) - B_1]B_7^t + e_t
\]

utilizing the (exact) long-run time trend term \((B_6^*t)\) resulting from the recursive transformation as derived in Fischer (2002), Appendix B, p. 21, see Annex B, below, top of p.11. Naturally, this expression reduces to

\[
\ln(iec_t) = B_1 + B_3^*\ln P_t + B_4^*\ln inv_{1t} + B_5^*\ln inv_{1-1} + B_6^*t
\]

\[+ B_2^*\ln(iec_{t-1}) + B_3^*\ln(iec_{t-2}) + B_7^2*\ln(iec_{t-3}) + \ldots + B_7^{t-1}*(\ln(iec_{0}))
\]

\[+ [\ln(iec_{0}) - B_1 - B_3^*\ln P_0 - B_4^*\ln inv_{0} - B_5^*\ln inv_{0} - B_6^*\ln P_0]*B_7^t + e_t
\]

reflecting a state adjustment framework (estimated over \(t=1=1967\) to \(t=31=1997\)) that is conceptually like the adaptive expectations model but with the infinite distributed lag (truncated at year \(0 = 1966\), in accordance with the availability of data) concentrated on the lagged dependent variable, thus indirectly generating a uniform rate of lagged adjustment with respect to the explanatory variables (as in models of the partial adjustment type). The weights on this series of lagged dependent variables decline geometrically at a rate of \(1-B_7\). The point estimates of the model’s parameters can be produced via LS by searching over the range of possible \(B_7\) values to select the regression that yields the lowest sum of squared residuals, but, since \(B_7\) is an unknown parameter, this approach does not provide proper estimates of parameter standard errors. Analyzing the model as if \(B_7\) is known precisely, it can be seen that LS estimates of parameters such as the \(B_3\) parameter on the associated composite lagged dependent variable are consistent but biased under conditions of limited data availability. This is because, while the composite lagged dependent variable and the disturbance \(e_t\) are uncorrelated, the former’s deviation from its sample mean entails \(\ln(iec_t)\) and is therefore not independent of \(e_t\), causing the expected value, as opposed to the probability limit, of estimated parameters to diverge from their true values. Similarly, LS parameter estimates for models including multiple lagged dependent variables with unrelated parameters are also consistent but biased in limited samples [e.g. Johnston (1984), ch.9]. Focusing on the full estimation of the infinite distributed lag model, each of the lagged values of \(\ln(iec_t)\) is uncorrelated with \(e_t\), but after demeaning, not independent of \(e_t\). With limited availability of data, this feature of each of the lagged dependent variables in the series can be expected to pose a potential challenge to nonlinear least squares maximum likelihood parameter estimation.

In all specifications based on the (first-order moving average) Klein/Zellner-Geisel method, the challenges of identifying significant variables can be simplified by avoiding the estimation of the (nuisance) parameter for the truncation remainder (which cannot be estimated consistently anyway), \([\ln(iec_{0}) - B_1 - B_3^*\ln P_0 - B_4^*\ln inv_{0} - B_5^*\ln inv_{0} - B_6^*\ln P_0]*B_7^t + e_t\).

equal to the product of $B_2$ and the initial value parameter, lagged state variable for year -1, (reducing the total number of parameters and increasing the estimate precision for those that remain), if the approximation that $E ln(iec_0) = ln(iec_0)$ [Pesaran(1973)] is deemed to be acceptable. Another simplification is to ignore the truncation remainder term entirely but past research experience [e.g. Dhrymes (1971); Schmidt (1975); Maddala and Rao (1971)] and that of Fischer (2002) suggests the omission of this term might compromise the finite sample efficiency of parameter estimates and, in particular, reduce the significance of the $B_7$ parameter (using the approximate long-run time trend term, $B_6*(1-B_7^t)^t$, as derived in Appendix B, instead of the exact term, might be marginally advantageous). However, this distinction between finite sample and asymptotic efficiency would disappear if the initial year of data availability were also the year (say year 0 = 1966) in which electricity service was initially extended to the region under study. If this were the case, the above (exact) truncation remainder, representing the infinite series $B_3*ln(iec_{-1}) + B_7*ln(iec_{-2}) + B_7^2*ln(iec_{-3}) + B_7^3*ln(iec_{-4}) + \ldots \ldots \ldots,$ would reduce to zero (setting iec prior to year 0 at the negligibly low value of one), so that the model would take the following simple form even in finite samples:

$$ln(iec_t) = B_1 + B_3*lnP + B_5*lnvai + B_5*protecto + B_6*t$$
$$+ B_7*[ln(iec_{-1}) + B_7*ln(iec_{-2}) + B_7^2*ln(iec_{-3}) + \ldots \ldots + B_7^{t-1}*ln(iec_{0})] + e_t$$

Overall, for this direct transformation Klein/Zellner-Geisel approach (given the modeling experience/data availability described in Fischer (2002)) we can anticipate the possibility that the series of lagged dependent variables may complicate prospects for maximum likelihood parameter estimation via nonlinear least squares. In this respect, modeling complications are even greater in the case of higher-order moving average structures, including the need to estimate a state adjustment and initial value parameter (the above initial value approximation is not applicable: see Annex A) for each state adjustment effect, recursively generating the lagged state variable from its initial values and lagged consumption. (Direct maximum likelihood estimation of ARMA models given higher-order moving average errors would entail maximizing complex log-likelihood functions with extra terms, an unpromising endeavor given the available data, as suggested earlier.)

(ii) Rough Equivalents or Approximations of Original ARMA Model for Maximum Likelihood Estimation via Nonlinear Least Squares

Second, as an initial alternative to the specifications of the previous section including the series of lagged dependent variables, a simple specification, shown here for $t=1$ (or $t=2$) to $t=31=1997$, features a uniform distributed lag on the differenced explanatory variables ($lnP$, $lnvai$, $protecto$, & time trend $t$) and constant to roughly reflect state adjustment dynamics, along with a well-behaved disturbance $e_t$. The uniform geometric distributed lag has parameter $B$, where $B$ represents the combined effect of the lagged state variable and first-order state adjustment, and is distinguished from the coefficient on the lagged dependent variable, $B_2 + B_7$, in the original ARMA specification shown above, as well as from the combined effect of $B_2$ and $B_7$ in the previous direct transformation of the original specification applying the Klein/Zellner-Geisel method and from $B_{2+7}$ in the next section. Utilizing all the lagged values available given the data, that is, taking $q=t$ for all $t$: 
\[ \ln(\text{iec}_t) = A_1(q) + B_3^*(\ln P_t - B_7^*\ln P_{t-1}) + \mathbf{B}^*(\ln P_{t-1} - B_7^*\ln P_{t-2}) + \mathbf{B}^2*(\ln P_{t-2} - B_7^*\ln P_{t-3}) + \ldots + \mathbf{B}^{q-1}*(\ln P_{t-(q-1)} - B_7^*\ln P_{t-q}) + B_4^*[(\ln\text{vai}_t - B_7^*\ln\text{vai}_{t-1}) + \mathbf{B}^*(\ln\text{vai}_{t-1} - B_7^*\ln\text{vai}_{t-2}) + \mathbf{B}^2*(\ln\text{vai}_{t-2} - B_7^*\ln\text{vai}_{t-3}) + \ldots + \mathbf{B}^{q-1}*(\ln\text{vai}_{t-(q-1)} - B_7^*\ln\text{vai}_{t-q})] + B_5^*[(\text{prodest}_t - B_7^*\text{prodest}_{t-1}) + \mathbf{B}^*(\text{prodest}_{t-1} - B_7^*\text{prodest}_{t-2}) + \mathbf{B}^2*(\text{prodest}_{t-2} - B_7^*\text{prodest}_{t-3}) + \ldots + \mathbf{B}^{q-1}*(\text{prodest}_{t-(q-1)} - B_7^*\text{prodest}_{t-q})] + A_6(q)^*t + \varepsilon_t \]

where \( A_1(q) = B_1^*(1-B_7^*)(1-B^2)/(1-B) \) and, for (an approximation to) the coefficient on the time trend [Annex B; Fischer (2002), Appendix B], \( A_6(q) = B_6^*(1-B_7^*)(1-B^3)/(1-B) \) (asymptotically, these expressions could be replaced by \( A_1 \) and \( A_6 \) and estimated as constants). This is a logical representation of demand dynamics that can be compared to those of the original ARMA specification of the state adjustment model as long as \( B \) significantly exceeds \( B_7 \), that is, the parameter \( B_2 \) on the lagged state variable is significantly greater than zero, at which value this representation degenerates into static form (given the nature of electricity consumption, as explained above, \( B_2 \) is not expected to be negative).\(^3\) In the same way, the representation can be compared to the specifications presented in section (i), for example, the above term featuring a distributed lag on the differenced \( \ln P \) variable is equal (where \( B_2 = B - B_7 \) to

\[ B_3^*\ln P_t + B_2^*(\ln P_{t-1} + B_3^*\ln P_{t-2} + \mathbf{B}^2*\ln P_{t-3} + \ldots + \mathbf{B}^{q-2}*\ln P_{t-(q-1)}) - B_7^*\mathbf{B}^{q-1}*\ln P_{t-q}], \]

showing the way the finite distributed lag scheme for \( \ln P \), \( \ln\text{vai} \), and \( \text{prodest} \) [for \( t \geq 2 \), this entails a weighted \((B_2/B; B_7/B)\) average of the \( t \)-1 to \( t-(q-1) \) and the \(-t-(q) \) lags, respectively], combined with that for the constant and time trend \([A_1(q)\) and \( A_6(q)^*t] \), can roughly reflect the demand dynamics of the infinite distributed lag scheme on the lagged dependent variable: Annex C justifies the related model that completes this connection.

It might not make much difference if the above equation is set up cutting off the distributed lags after a reduced number of periods \( s \) for the differenced variables \( \ln P \), \( \ln\text{vai} \), and \( \text{prodest} \) and, correspondingly, for the differenced constant and time trend, in that case \( q=t \) for \( t \leq s \) and \( q=s \) for \( t > s \), but at least \( s \) should remain quite large.

---

\(^3\) For a long-run equilibrium relationship among the variables to exist in the formal state adjustment first-order moving average electricity demand framework, it is necessary for the inequality \( B_7 + B_2 < 1 \) and thus \( B_7 < 1 - B_2 \) to be satisfied: assuming the variables are integrated of order 1, \( I(1) \), the original ARMA model re-expressed in error correction form is

\[ \Delta\ln(\text{iec}_t) = B_1^*(1-B_7^*) + B_6^*(1-B^2)^*(t) + B_3^*\Delta\ln P_t + B_4^*\Delta\ln\text{vai}_t + B_5^*\Delta\text{prodest}_t - B_3^*[(\ln\text{iec}_{t-1}) - B_7^*(1-B_3^*/B^2)^*\ln P_{t-1} - B_7^*[(1-B_3^*/B^2)^*\ln\text{vai}_{t-1} - B_7^*[(1-B_3^*/B^2)^*\text{prodest}_{t-1}]) + \varepsilon_t - B_7^*\varepsilon_{t-1} \]
an (a comparable formulation involves setting \( B_2 \) to zero), where \( -B_3^* = (B_7 + B_2 - 1) \) is the coefficient on the deviation from the long-term relationship, which should be significantly less than zero. Alternatively, assuming the variables are \( I(1) \) after quasi-first-differencing, in the case of the special autocorrelation scheme of section (iii), produces an analogous error correction expression, with a well-behaved disturbance, the same coefficient on the deviation from the transformed long-term relationship, and an equilibrium error with respect to the untransformed long-term relationship that is a second-order, as opposed to a first-order, autoregressive process. The inequality \( B_7 + B_2 < 1 \) can also serve as the standard for judging the consistency of the specifications that are rough equivalents or approximations of the original ARMA model with the economic implications, assessed in chapter 1, of the Houthakker and Taylor (1966, 1970) state adjustment demand theory.
(iii) Original ARMA Model Moving Average Disturbance Term Subject to Autocorrelation – Maximum Likelihood Estimation via Nonlinear Least Squares

Finally, another interesting alternative approach to the estimation of the state adjustment model [consistent with the conceptual cases suggested by Carter and Zellner (2003)], one that could be particularly apropos for handling the associated moving average disturbance term (though it does entail working with lagged dependent variables for periods t-1 and t-2) is based on the special hypothesis that this disturbance term is itself subject to first-order autocorrelation. A first-order autocorrelation transformation of the original ARMA specification results in a model with a well-behaved disturbance, as follows:

\[
\ln(\text{iec}_t) - B_8\ln(\text{iec}_{t-1}) = B_1 + B_3[\ln(P_t - B_7\ln(P_{t-1}) - B_8[\ln(P_{t-1} - B_7\ln(P_{t-2})] \\
+ B_4[\ln(\text{inva}_{t} - B_7\ln(\text{inva}_{t-1}) - B_8[\ln(\text{inva}_{t-1} - B_7\ln(\text{inva}_{t-2})] \\
+ B_5[(\text{protecto}_{t} - B_7\text{protecto}_{t-1}) - B_8(\text{protecto}_{t-1} - B_7\text{protecto}_{t-2})] + B_6t \\
+ (B_{2+7})*[\ln(\text{iec}_{t-1}) - B_8\ln(\text{iec}_{t-2})] + w_t
\]


where \((e_t - B_7^*e_{t-1}) = B_8^*(e_{t-1} - B_7^*e_{t-2}) + w_t\) (\(w_t\) normally and independently distributed); \(B_8\) is the parameter of the AR(1) scheme on \((e_t - B_7^*e_{t-1})\); \(B_{2+7} = B_2 + B_7\); \(B_1 = B_1^*(1-B_7)*(1-B_8)\); and \(B_6 = B_6^*(1-B_7)*(1-B_8)\).

Here it is straightforward to define similar models for second- or higher-order moving average state adjustment structures on the basis of the hypotheses that their corresponding moving average disturbance terms are subject to first-order autocorrelation. Besides the autocorrelation parameter, only one additional (state adjustment) parameter needs to be estimated for each higher-order state adjustment effect (there are no initial value parameters to be estimated). In the case of an AR(1) hypothesis with regard to a second-order moving average disturbance term \((e_t - B_7^*e_{t-1} - C_7^*e_{t-2})\), where \(C_7\) is the second-order state adjustment parameter, for each explanatory variable \(X\) (\(\ln(P, \text{inva}, \text{or protecto})\)), the term \((X_t - B_7^*X_{t-1} - C_7^*X_{t-2})\) replaces \((X_t - B_7^*X_{t-1})\), likewise for the associated lagged autocorrelation-correction term, a second-order lagged dependent variable term \(C_7^*[\ln(\text{iec}_{t-2}) - B_8^*\ln(\text{iec}_{t-3})]\) is added to the equation, \(B_1\) becomes \(B_1^*(1-B_7-C_7)*(1-B_8)\), and \(B_6\) becomes \(B_6^*(1-B_7-C_7)*(1-B_8)\).

Even in the first-order moving average case, this additional alternative specification is more complex than the specifications successfully estimated in Fischer (2002). Given this past modeling experience in the context of limited data availability, considerable difficulties can also be expected with maximum likelihood parameter estimation via nonlinear least squares (over \(t=2=1968\) to \(t=31=1997\), conditional on \(\ln(\text{iec}_1)\)) in this case.

The alternative hypothesis of a first-order autoregressive scheme on \(e_t\) in the specifications derived in section (i) above (which feature the series of lagged dependent variables) produces less plausible transformed specifications of greater complexity. Likewise for \(e_t\) in the specifications developed in section (ii) and its annexes.
Conclusion

The relatively straightforward maximum likelihood estimation method utilizing nonlinear least squares can be applied to the alternative specifications whose conceptual interrelationship is explored in this note and its annexes. These specifications are substantially more complicated than the specifications successfully estimated in Fischer (2002). On the basis of this past modeling experience and the corresponding limited availability of data, we can still anticipate considerable parameter estimation difficulties (even in the case of the alternative specifications derived from the simplest, first-order moving average forms of state adjustment dynamic demand modeling), in particular, for the $B_2$, $B_7$, $B_8$ and $B_5$ parameters. However, these specifications are worth re-examining on the basis of future expanded (still limited) data availability.

Acknowledgments

Comments on earlier drafts of this paper encouraged additional explicit analytical elaboration of the way the model conceived, specified, and justified at the end of Annex C approximates the theoretical foundations of the state adjustment model.
References


ANNEX A

Section (i) Approximation of Initial Values

In demonstrating the way in which, in section (i), the initial value approximation is applicable to the first-order moving average state adjustment model, and not to higher-order models, this annex highlights the relevant characteristics of the structure of the corresponding distributed lag schemes.

For the first-order moving average case, the expected value of electricity consumption in year 0 = 1966, the initial year of data availability, is

\[ \text{Eln}(\text{iec}_0) = B_1 + B_3 \cdot \ln P_0 + B_4 \cdot \ln v_0 + B_5 \cdot \text{protecto}_0 + B_2 \cdot S_{-1} \]

where \( S_{-1} \) is the lagged state variable for year -1. Going backwards in time, \( S_{-1} \) is itself the product of the previous lagged state variable in first-order state adjustment scheme \( S_{-1} = \ln(\text{iec} - 1) + B_3 \cdot S_{-2} \), and so on, and is thus a function of an infinite series of unknown annual electricity consumption data points going back from the year prior to year 0:

\[ S_{-1} = \ln(\text{iec} - 1) + B_3 \cdot \ln(\text{iec} - 2) + B_3^2 \cdot \ln(\text{iec} - 3) + B_3^3 \cdot \ln(\text{iec} - 4) + \ldots \ldots \]

In the infinite distributed lag specification of section (i), the (constant) truncation remainder parameter is the expected value of electricity consumption in year 0 minus the constant term and the explanatory variable initial (year 0) data terms. Approximating the truncation remainder by setting \( \text{Eln}(\text{iec}_0) = \ln(\text{iec}_0) \) makes sense given the understanding that it is theoretically possible to predict its value on the basis of the product of \( B_2 \) and the \( S_{-1} \) function. Because of this function’s simple, first-order progression of coefficients, putting the value of the truncation remainder at the actual year 0 level of electricity consumption, corrected for the constant and the explanatory variable initial data terms, may well be viewed, in principle, as a plausible approximation, conceptually consistent with the propagation of this first-order state adjustment pattern into the future.

To compare the higher-order moving average case on equal terms, the expected value of electricity consumption, also in year 0, is

\[ \text{Eln}(\text{iec}_0) = B_1 + B_3 \cdot \ln P_0 + B_4 \cdot \ln v_0 + B_5 \cdot \text{protecto}_0 + B_2 \cdot S_{-1} \]

where \( S_{-1} \) is the lagged state variable for year -1 for the higher-order case. Going backwards in time, in the second-order model, \( S_{-1} \) is itself the product of the second-order state adjustment scheme \( S_{-1} = \ln(\text{iec} - 1) + B_3 \cdot S_{-2} + B_4 \cdot S_{-3} + (B_2 + C_7) \cdot S_{-4} \) (\( C_7 = \) second-order state adjustment parameter) and depends on both \( S_{-2} \) and \( S_{-3} \). In the next stage, this expands to

\[ S_{-1} = \ln(\text{iec} - 1) + B_3 \cdot \ln(\text{iec} - 2) + B_3^2 \cdot S_{-3} + B_7 \cdot B_3 \cdot S_{-4} + C_7 \cdot S_{-5} + \ldots \ldots \]

The expanding web plays out indefinitively into the past and, after collecting terms for each year of unknown electricity consumption data points prior to year 0, the result is an infinite function that looks like:
\[ S_{1} = \ln(\text{iec}_1) + B_7 \ln(\text{iec}_2) + (B_7^2 + C_7) \ln(\text{iec}_3) + (B_7^3 + 2B_7C_7) \ln(\text{iec}_4) \]
\[ + (B_7^4 + 3B_7^2C_7 + C_7^2) \ln(\text{iec}_5) + (B_7^5 + 4B_7^3C_7 + 3B_7^2C_7^2) \ln(\text{iec}_6) \]
\[ + (B_7^6 + 5B_7^4C_7 + 6B_7^2C_7^2 + C_7^3) \ln(\text{iec}_7) \]
\[ + (B_7^7 + 6B_7^5C_7 + 10B_7^3C_7^2 + 4B_7^2C_7^3) \ln(\text{iec}_8) \]
\[ + (B_7^8 + 7B_7^6C_7 + 15B_7^4C_7^2 + 10B_7^2C_7^3 + C_7^4) \ln(\text{iec}_9) + \ldots \ldots \]

The higher-order complexity of the progression of coefficients of the \( S_{1} \) function (multiplied by \( B_2 \)) makes it clear that a simple, first-order approximation, like the one applied in the above case, is not conceptually applicable to this higher-order process. Specifically, it does not reasonably follow on the basis of the theoretical prediction of this complex function, to fix \( B_2 \) at the value of the above approximation of the (first-order) truncation remainder. This would require ignoring higher-order terms, i.e., taking \( C_7 \) and the influence of higher-order effects as very small, in which case it might well be better to consider returning to the less complicated problem of estimating the first-order model. In the higher-order case, what is called for is the estimation of a number of initial value parameters equal to the order of the state adjustment model [for the rational lag structure: Dhrymes, Klein and Steiglitz (1970); Maddala and Rao (1971)]. However, for further clarity, the idea of integrating the above approximation into the framework of a higher-order state adjustment estimating equation, to save on the total number of parameters, is examined below. With the second-order state adjustment model, a conceptually sound approach, given the initial known data point for \( \ln(\text{iec}_0) \), is to utilize two lagged state variables, \( S_{1} \) and \( S_{2} \), as initial value parameters, \( S'_{-1} \) and \( S'_{-2} \), in the estimating equation

\[ \ln(\text{iec}_t) = B_1 + B_3 \ln P_t + B_4 \ln \text{vai}_t + B_5 \text{protec}_t + B_6 t + B_2 S_{t-1} + e_t \quad (t=1 \text{ to } t=31) \]

in order to recursively generate the lagged state variable \( S_{t-1} \) forward in time from

\[ S_0 = \ln(\text{iec}_0) + B_7 S'_{-1} + C_7 S'_{-2}. \]

Of course, the complex progression of coefficients applying to the series of lagged dependent variables [starting with \( \ln(\text{iec}_0) \)] that enters the estimating equation in this way is the same progression of coefficients that plays out (top of this page) as a result of the backward expansion elaborated in this annex. In addition, starting with \( B_7 \), this same progression of terms comprises the regression data series corresponding to the \( S'_{-1} \) parameter. Starting with \( C_7 \), a distinct but interrelated progression of terms, which disappears if and as \( C_7 \) gets small, comprises the data series for the \( S'_{-2} \) parameter:

\[ [(C_7), (B_7^2 C_7), (C_7^2 + B_7^2 C_7), (2B_7 C_7 + B_7^2 C_7), (C_7^3 + 3B_7^2 C_7^2 + B_7 C_7), \]
\[ (4B_7^3 C_7^2 + 3B_7^2 C_7^3 + B_7 C_7), (C_7^4 + 5B_7^3 C_7^2 + 6B_7^2 C_7^3 + B_7 C_7), \]
\[ (6B_7^4 C_7^2 + 4B_7^3 C_7^3 + 10B_7^2 C_7^4 + B_7 C_7), \ldots \ldots \ldots]. \]

Approximating \( B_2 S'_{-1} \) by substituting \([\ln(\text{iec}_0) - B_1 - B_3 \ln P_0 - B_4 \ln \text{vai}_0 - B_5 \text{protec}_0] \) might substantially alter model parameter estimates and negatively impact finite sample efficiency if a real higher-order state adjustment process is at work, i.e. if \( C_7 \) is not in fact very small, thus compounding the problem posed by series of lagged dependent variables.
ANNEX B

Specification of Section (ii): Full Derivations of the Coefficient on the Time Trend

The time trend term $B_6^t$ of the basic structural equation of the Houthakker-Taylor state adjustment model takes the following full form in the original autoregressive moving average (ARMA) modeling approach discussed at the beginning of this note:

$$B_6^t[(t-t-1)] + (1-B_7^t)B_6^t(t-1) = B_6^t[t - [B_7^*[t*(t-1)]]] = B_6^t(1-B_7^t)[t + [B_7^*(1-B_7^t)]]

The constant $+B_7^*/(1-B_7^t)$ is not shown in the original ARMA specification identified on p.1: it cancels out of the (exact) time trend term in the direct transformation of this model in section (i) (p.3) and has no effect on the time trend’s period count in the special autocorrelation transformation of this model in section (iii) (p.6). However, it does not cancel out of the specifications of section (ii), which leads to the need to rely on an approximation to the exact time trend expression (p.5), as explained in this annex.

Expressing the above full form of the original ARMA model’s time trend term as a finite distributed lag for the purposes of the distributed lag specifications of section (ii), and, initially, utilizing all the lagged values given the data (that is, taking $q$ as equal to $t$ for all $t$), we have:

$$B_6^t(1-B_7^t)[t + B^*(t-1) + B^2*(t-2) + \ldots + B^{t-1}*[t-(t-1)] + [B_7^*/(1-B_7^t)][1+B+B^2+\ldots+B^{t-1})]

A sequence of reformulations that are analogous to the steps in the recursive transformation of the time trend term of a set of autoregressive partial adjustment and adaptive expectations models, laid out on p.21 of Fischer (2002) (in Appendix B), derives the following exact expression:

$$[B_6^t(1-B_7^t)/(1-B^t)]*[(1-B^t)^*t - [B/(1-B) - B/(1-B^t)]*(1-B^t) + (B^t)*t]

Since $B \neq B_7^t$, the constant $- [B/(1-B) - B_7/(1-B_7^t)]$ in this expression does not simply cancel out as in Fischer (2002), bottom of p.21, top of p.22. This expression therefore features both the product of $-B^t$ and the sum of $t$ and this constant, and the product of $B^t$ and $t$. Thus, this exact time trend expression remains rather unwieldy for purposes of estimation.

However, it is possible to produce a good approximation to this exact expression following the assumptions developed in Fischer (2002), Appendix B. The approximation shown there actually improved upon the finite sample parameter estimation results associated with that exact expression, which took the simplest of forms. Rearranging this expression yields:

$$[B_6^t(1-B_7^t)(1-B^t)/(1-B)]*t - [B/(1-B) - B_7/(1-B_7)] + [(B^t)/(1-B^t)]*t}
Now, for plausible values of $B$, the term $[(B^i)/(1-B^i)]^t$ can be seen as fading into insignificance at low values of $t$. For example, for $B$ as high as 0.6, at $t$ as low as 8 the term has already lost 90.9% of its value (at $t$ equal to 1). In this situation, the terms within the braces of the last expression can be taken as effectively equivalent to $t$ minus a constant that does not vary with time. Since a constant leaves the time trend’s period count unchanged, it is therefore dropped from the expression.

In contrast, the term $(B^i)^t$ (also when it represents the product of $B^i$ and the period count) does not fade away quite as fast. Only if it did, could one expect that expressing the above exact expression as $[B_6*(1-B_7)/(1-B)]^t$ would be an equally good approximation. Thus, the best approximation of the time trend expression can be expected to be:

$$A\delta(q)^t = [B_6*(1-B_7)*t]/(1-B)]^t$$

The derivation of a practical time trend expression for the case in which the finite distributed lag scheme on the above original ARMA model full form is cut off after a number of periods $s$ that falls short of the number of periods corresponding to the availability of data (that is, $q=t$ for $t \leq s$ and $q=s$ for $t > s$), proceeds in a manner that is subtly different. The second expression on the previous page must now be evaluated as follows:

$$B_6*(1-B^7)*\{t + B^*(t-1) + B^2*(t-2) + ... + B^{q-1}*[t-(q-1)] + [B_7/(1-B_7)](1+B+B^2+...+B^{q-1})\}$$

For this case, the actual sequence of reformulations that is analogous to the derivation in Fischer (2002), p.21, is laid out in the next four steps below:

$$B_6*(1-B^7)*\{t*[1+B+B^2+B^3+...+B^{q-1}] - [B+2*B^2+3*B^3+...+(q-1)*B^{q-1}] + [B_7/(1-B_7)](1+B+B^2+...+B^{q-1})\}$$

Reformulating and adding and subtracting $-B*[q*B^{q-1}]*B_6*(1-B_7)$ produces:

$$B_6*(1-B^7)*\{t*[1+B+B^2+B^3+...+B^{q-1}] - [B+2*B^2+3*B^3+...+(q-1)*B^{q-1} + q*B^q + q*[B^q] + [B_7/(1-B_7)](1+B+B^2+...+B^{q-1})\}$$

After reformulating the series $[1+2*B+3*B^2+...+q*B^{q-1}]$ (i.e., multiplying it by $B$, subtracting the result from it, and dividing by $1-B$), the expression becomes:

$$B_6*(1-B^7)*\{t*[1+B+B^2+B^3+...+B^{q-1} - q*B^q] + q*[B^q] + [B_7/(1-B_7)](1+B+B^2+...+B^{q-1})\}$$

Reformulating and simplifying leads to the following exact expression in this case:

$$[B_6*(1-B^7)/(1-B)]*[t - (B/(1-B) - B_7/(1-B_7))] + (B^q)*q$$
Again, this is an unwieldy expression for purposes of estimation: \( B \neq B_7 \), and also, for \( q=s \), \((B^q)*t \neq (B^q)*q \), such that when \( t > s \) the expression is:

\[
[B_6(1-B_7)/(1-B)]*\left\{ \left[ (1-B^s)*(1-(s/t)) \right] - \left[ B/(1-B) - B_7/(1-B_7) \right] \right\} * (1-B^q)
\]

A better way to proceed is to further develop the exact expression (last expression on the previous page) to derive an approximate expression from what in this case looks like the following:

\[
[B_6(1-B_7)*(1-B^q)/(1-B)]*\left\{ t - \left[ B/(1-B) - B_7/(1-B_7) \right] + \left[ (B^q)/(1-B^q) \right] * q \right\}
\]

As long as the number of periods \( s \) is not too far below the number of distributed lags that are available given the data, it is possible for the third term within the braces of this expression to fade into insignificance (while \( q \) is still rising below the limit \( s \)) after only a small number of periods \( t \). If such a situation holds (depending, of course, on the value of \( B \), for example, plausible values as high as the range considered on the top of p.12), the terms within the braces can be viewed as effectively equivalent to \( t \) minus a constant. [In contrast, the term \((B^q)*t \) (also when it represents the product of \( B^q \) and the period count) --- and, likewise, the term \((B^q)*q \), see the exact expression --- do not fade away quite as fast.] Therefore, dropping the constant, the time trend expression is approximated as:

\[
A_6(q)*t = [B_6(1-B_7)*(1-B^q)/(1-B)]*t \quad \text{where } q=t \text{ for } t \leq s \text{ and } q=s \text{ for } t > s.
\]

To carry out the estimation, starting from \( t=2 \), in the finite distributed lag framework of section (ii) there are two possible approaches. The simpler and more consistent approach is to utilize the simple specification on p.5 (including the above derivations) as is, that is, take the distributed lag scheme on the differenced explanatory variables, including time trend, and constant as beginning in period 1 and not fully developed until period 2. The other possible approach is to assume the distributed lag scheme is initiated with the first observation of period 2 for the differenced constant and time trend even though this puts these terms out of phase with the other variable lags, which again reflect the development of the scheme that is initiated in period 1. Then, the finite distributed lag scheme for the constant and the time trend covers \( p \) periods, where \( p = q - 1 \), with \( q \) again defined as in the above cases: \( q=t \) for all \( t \) or \( q=t \) for \( t \leq s \) and \( q=s \) for \( t > s \). The constant term becomes:

\[
A_1(p) = B_1(1-B_7)*(1-B^p)/(1-B), \text{ which is equivalent to } A_1(q) = B_1(1-B_7)*(1-B^{q-1})/(1-B)
\]

Similarly, the time trend term, based on the derivations of this annex, becomes:

\[
A_6(p)*t = [B_6(1-B_7)*(1-B^p)/(1-B)]*t \quad \text{or} \quad A_6(q)*t = [B_6(1-B_7)*(1-B^{q-1})/(1-B)]*t
\]

This approach, entailing a one-period shift in the constant and time trend terms, substantially alters the simple specification shown on p.5 of the text. In contrast, it is this simple specification of the finite distributed lag scheme that is apropos regarding the related, more complete, modeling framework described at the end of Annex C.
ANNEX C
Rebalancing the Model towards the Theory-Based Dynamic Structure of Section (i)

This annex delves deeper into the connection between the finite distributed lag scheme of section (ii) and the infinite distributed lag scheme of section (i). Because $B$ (equal to positive $B_2$ plus $B_7$) exceeds $B_7$, the finite distributed lag scheme on $\ln(P)$, $\ln(vai)$, protecto, the constant, and the time trend, as a proxy for the infinite distributed lag scheme on lagged electricity consumption, puts greater relative and absolute weight on nearer as opposed to more distant lags. While weighing periods $t$ and $t-1$ equally, the series

$$B_3[\ln(P_t) + B_2(\ln(P_{t-1}) + \frac{B\ln(P_{t-2}) + B^2\ln(P_{t-3}) + \ldots + B^{t-2}\ln(P_1)}{1-B^t})] - \ldots$$

from section (ii) puts greater relative and absolute weight on $\ln(P_{t-2})$ through $\ln(P_1)$ than the series

$$B_3[\ln(P_t) + \ldots + B_2[\ln(\text{iec}_{t-1}) + B_2\ln(\text{iec}_{t-2}) + B_7\ln(\text{iec}_{t-3}) + \ldots + B_7^{t-2}\ln(\text{iec}_1)] + \ldots$$

from section (i) puts on $\ln(\text{iec}_{t-2})$ through $\ln(\text{iec}_1)$.

The same holds for $\ln(vai)$ and protecto. In addition, the constant and (approximate) time trend terms in section (ii) are greater than the corresponding terms in section (i) by a factor of $(1-B)/(1-B)$. In both cases, the initial values of these terms, along with the explanatory variable initial terms for year 0, reflect the structure of state adjustment, but the above section (i) series includes lagged dependent variable $\ln(\text{iec}_0)$, explaining $\ln(\text{iec}_1)$, with coefficient $B_2B_7^{-1}$ (no counterpart in the section (ii) series), while also serving to connect with the long-run equilibrium adjustment mechanism. In sum, the section (ii) scheme dispenses completely with the econometrically complicating series of lagged dependent variables and shifts the weight accorded to distant lags of the explanatory variables in the infinite distributed lag section (i) scheme onto the nearer lags for which data is available, conceivably picking up lagged impacts more realistically.

Now, adding the initial value of consumption term $\ln(\text{iec}_0)\frac{B^t}{1-B}$ to the initial value elements of the simple specification of section (ii) has the effect of shifting this weight back to a partial extent by creating a proxy truncation remainder term that is similar to the corresponding approximate truncation remainder term of section (i), p.4, but with a proxy truncation remainder, in brackets, that is smaller, as indicated next:

$$[\ln(\text{iec}_0) - (B_1 + B_6^*t)(1-B)/(1-B) - B_2^*(B_7/B)*\ln(P_0) - B_4^*(B_7/B)*\ln(vai_0) - B_5^*(B_7/B)*\text{protecto}_0]^*B^t$$

The model can be estimated using the methods of this note with the assumption that the disturbance is well-behaved [starting at $t=1$, where there is no distinction (canceling $B$) from the model of section (i), p.4]. It is a model that can be derived from the original ARMA state adjustment model through recursive transformation to eliminate the lagged dependent variable $\ln(\text{iec}_{t-1})$, but not with a well-behaved disturbance. The $\ln(\text{iec}_0)^*B^t$ term reduces the weights of the simple specification’s (near) lags, with $B_2/B$ falling; the proxy truncation remainder term is closer to the approximate truncation remainder term of the grounded-in-theory infinite distributed lag scheme, the closer $B$ gets to $B_7$ (ftn1).