Moving Averages? Moving Averages.

Garrett N. Vande Kamp*

Soren Jordan †

Abstract: The debate over lagged dependent variables is over. By including enough lags of the dependent and independent variables, it is thought that scholars can fully model autocorrelation in the disturbance term and consistently estimate short-run and long-run marginal effects. We illustrate that this approach fails to account for something we've known about for a long time but have since forgotten: moving averages. The failure to incorporate moving averages into this modelling strategy is problematic, as moving averages cannot be represented with a finite number of lags. Failure to model this class of autocorrelation in the disturbances will lead to either inconsistent or inefficient estimates of relevant quantities of interest. But the inclusion of moving averages into this modelling strategy is both theoretically and computationally feasible and results in a dynamically complete model that is equivalent to a transfer function model. These claims are demonstrated through a set of Monte Carlo simulations.

*University of Georgia; garrettvandekamp@uga.edu

†Auburn University; scj0014@auburn.edu

1 Introduction

After almost two decades of debate, political science has finally settled the debate over lagged dependent variables in dynamic regression models. After a concerning article by Achen (2000) demonstrating bias when a model with a lagged dependent variable was estimated when autocorrelation remained in the disturbance term, scholars often avoided estimating such models. Soon after, Keele and Kelly (2006) reminded authors that deliberately omitted a lagged dependent variable with explanatory power would result in omitted variable bias. Scholars were left to struggle with these competing concerns until Wilkins (2018) demonstrated that autocorrelation in the disturbance term could be fully modelled with additional lags of the dependent and independent variables. Indeed, practitioners are currently advised to add more lags of variables until all of the autocorrelation in the disturbance term is absorbed (particularly common in models from the econometric perspective of error as a nuisance, like Philips (2018)).

Yet throughout the prior saga, authors made one crucial assumption about the data-generating process. The autocorrelation in the error term resulted from an autoregressive error process. While such a process is certainly possible in the real world, it only represents one of two possible sources of autocorrelation in the error term. The other, moving averages, is largely omitted in this discussion. Indeed, moving averages have largely been omitted from dynamic analysis in political science, despite being covered in virtually every graduate seminar on the subject.

We argue that excluding moving averages from dynamic regression models poses a grave risk to political scientists. Much like with autoregressive errors, failure to attempt to model a moving average will result in significant bias in important quantities of interest. Unlike with autoregressive errors, however, completely modelling a moving average term with lagged dependent and independent variables is impossible. *Approximating* these processes may be possible, but doing so will result in inefficiencies compared to explicitly modelling a moving average. Therefore, we advocate that scholars specify dynamic regression models that include moving averages.

We begin with a review over the debate over lagged dependent variables to ground the reader. We next reintroduce the moving average to political science, paying particular attention to why moving averages have largely been forgotten in applied time series analysis. We then demonstrate the perils of omitting moving averages from dynamic regression models through Monte Carlo simulations. We then describe how dynamic models with moving averages can be estimated, interpreted, and incorporated into existing strategies for modelling a time series. We conclude with our thoughts on what these results say for other fields in political science.

2 Autocorrelated Errors and The Debate Over Lagged Dependent Variables

The debate over lagged dependent variables in political science is, at its core, a debate about modelling autocorrelation in the disturbance term. Once viewed as a mere nuisance to be corrected, political scientists came to view autocorrelation in the disturbance term as an opportunity create and model dynamic theories in political science (Beck, 1985). The most common approach to transform this autocorrelation into a theoretically meaningful quantity was to include a lagged dependent variable in dynamic regressions (Keele and Kelly, 2006). While such a strategy is certainly effective, scholars noticed that doing so seemed to drain the explanatory power of the independent variables in the model.

Achen (2000) was the first to clearly demonstrate why this phenomena occurred in his now famous critique of the use of lagged dependent variables. As an example, he set up an ADL(1,0) data-generating process with autoregressive errors:

$$y_t = \alpha y_{t-1} + \beta x_t + u_t \tag{1}$$

$$x_t = \rho x_{t-1} + \varepsilon_{1t} \tag{2}$$

$$u_t = \phi u_{t-1} + \varepsilon_{2t} \tag{3}$$

where u_t is the disturbance of the dependent variable y_t and ε_1 and ε_2 are both white noise error terms.[1]

Achen demonstrates that when the autoregressive parameter of the error term, ϕ, is non-zero, estimates of α from an OLS regression of equation 1 will be biased. Further, if θ is also non-zero, estimates of β will also be biased and the bias in α will also grow. This put researchers in a difficult position. If they include a lagged dependent variable, they risk bias permeating throughout the model. If they exclude it, however, they risk omitted variable bias, which could also result in bias permeating throughout the model (Keele and Kelly, 2006). This led to many debates in the discipline over the used of lagged dependent variables.

(Wilkins, 2018) demonstrated that this debate was largely trivial because a properly specified regression model could accommodate the autocorrelation in equation **??** while still accurately estimating both the short-run and long-run effects of x on y. Utilizing the common factor restriction for models with autoregressive errors, he demonstrated that equations 1 and 3 could be combined into a single ADL (2,1) model of the following form:

$$y_t = (\alpha + \phi) y_{t-1} - \alpha\phi y_{t-2} + \beta x_t - \beta\phi x_{t-1} + u_t \tag{4}$$

[1] Or an unobserved, random variable that is not itself dynamically autocorrelated

From reading Wilkins' argument, one would assume that autocorrelation in the error term can be fully resolved by including enough lags of the dependent and independent variables. After all, Wilkins and his predecessors use the general term of autocorrelation to describe the dynamic process in equation 3. Yet Wilkins' strategy carries an important, yet unstated, caveat. Additional lags of the dependent and independent variables can model autocorrelation in the disturbance term if that autocorrelation is caused by autoregression. Yet autocorrelation in the error term is not limited to autoregression. Indeed, there are two major classes of univariate dynamic processes that result in autocorrelation, only one of which is autoregression.

3 Moving Averages

Autoregressive errors make an important assumption about the dynamic effect that the disturbance term has on the dependent variable. Namely, they assume that the disturbance term will always have some effect on the dependent variable, no matter how much time has passed. Stationarity requires this effect to decrease as the amount of time between two points increases, but, no matter how far in the past the disturbance term is, it will always have a non-zero effect on the present dependent variable. For example, the autoregressive error in equation 3 can be reexpressed as an infinitely decreasing function of white noise error terms:

$$u_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + ... \tag{5}$$

Moving averages are a second class of dynamic error terms. Whereas autoregression in the disturbance term results in the disturbance term being a function of lags of itself and a white noise process, moving averages in a disturbance term that is a function of a white noise error process and lags of that white noise process. A first-order moving average that is comparable to the first-order autoregressive error in equation 3 is written as:

$$u_t = \varepsilon_t + \theta\varepsilon_{t-1} \tag{6}$$

The distinction between the two is subtle, but important. In the moving average case, the dynamic effect of the error term is finite: after a certain point, previous iterations of the error term no longer influence the dependent variable. This is opposed to the infinite effect the error term has in autoregressive error processes. Moving averages are similar to lagged independent variables, which also have a dynamic effect that becomes zero after a certain point in time. Table 4 contains a handy comparison chart for the different potential components of a dynamic data-generating process.

Table 1: Potential Components of a Dynamic Data Generating Process, by Observability and Dynamic Effect

	Covariates	Error Term
No Dynamic Effect	Independent Variable	White Noise Error
Finite Dynamic Effect	Lagged Independent Variable	Moving Average
Infinite Dynamic Effect	Lagged Dependent Variable	Autoregressive Error

Theoretically, a moving average can be interpreted as an omitted variable that is not itself autocorrelated but, nonetheless, has a delayed effect on the dependent variable (Slutsky, 1937). Of course, not all moving averages have theoretical interpretation. For example, if an autoregressive process is measured with error, such error will result in a moving average process (Beck, 1985). Unlike autoregressive terms, moving averages do not threaten the stationarity of a series. This is due to non-uniqueness of moving averages, in which a moving average with a coefficient of greater than one can be equivalently reexpressed as a coefficient of less than one (Shumway and Stoffer, 2017, 91). For this reason, we will restrict our attention to moving averages with coefficients less than one without a substantial loss of generality.[2]

[2]The only case excluded is when a moving average has a coefficient equal to exactly one. This is called a non-invertible moving average. Non-invertible moving averages can appear in a model, such as when a

Moving averages and autoregressive parameters are directly related due to their invertibility. In a stationary series, autoregressive terms can be reexpressed as an infinite number of moving averages and lags of the independent variables. Similarly, moving averages can be reexpressed as an infinite number of autoregressive terms and lags of the independent variables. For example, if a data-generating process was characterized by equations 1 and 5, it could be reexpressed as:

$$y_t = (\theta + \alpha)y_{t-1} - \theta(\theta + \alpha)y_{t-2} + \theta^2(\theta + \alpha)y_{t-3} - ... + \beta x_t - \theta\beta x_{t-1} + \theta\beta x_{t-2} - ... + u_t \quad (7)$$

This reexpression of a moving average should cause readers to note a few things up front. Much like with autoregressive errors, a moving average is likely to induce correlation between the error term and both lagged dependent and independent variables. Unlike with autoregressive errors, however, this autocorrelation cannot be reexpressed using a finite number of lags of both lagged dependent and independent variables; an infinite number of lags is required. Thus, the only means to capture all of autocorrelation in the error term induced by moving averages is by explicitly including moving averages into a regression model.

While moving averages were first proposed in the early twentieth century, they were rarely used in practice due to the complexity of estimating models containing them. At present time, advances in both computing power and optimization routines make the estimation of models with moving averages trivial. Yet moving averages are not universally estimated in dynamic models either political science or other fields. To understand why, it is helpful to review the two most prominent methodologies for studying time series data -

stationary series is differenced. But given that the coefficient of a moving average can be any real number and thus follows a continuous probability distribution, the probability of encountering a true moving average with a coefficient equal to exactly one is zero.

the Box-Jenkins methodology and Hendry's general-to-specific methodology - and examine how moving averages processes are modelled in each.

4 Moving Averages and Box-Jenkins

In their seminal work, Box and Jenkins (1976) introduced the ARIMA model. After ensuring that a series was stationary (and differencing if necessary, due to a series being integrated), the analyst would reduce a univariate input process to white noise residuals through the use of autoregression and moving average terms, described above. Once this white-noise error process was properly filtered, the analyst could check for the explanatory role of external regressors through the use of transfer functions. These transfer function models, essential to the ARIMA strategy, shine in their flexibility. They can flexibly model finite and infinite dynamic effects of both independent variables and the error term; all of the terms in Table are captured by transfer function models. Though for reasons that will become apparent, the use of transfer functions and the Box-Jenkins methodology in political science are now relatively rare.

The first step to the Box-Jenkins methodology is ensuring a univariate series has white-noise residuals through the ARMA fitting processes (on a stationary series). Box and Jenkins viewed the two sides of the autoregressive moving average (ARMA) model as complementary, neither of which was *a priori* elevated above the other. [3] Just as much attention was given to MAs as to ARs: for example, for series which we might imagine the effect of shocks to be constrained to q periods, only the moving average could practically constrain these effects.

However, Box and Jenkins were limited by their contemporary place in computing and history. The full estimation of an ARMA model required an iterative cycle of building,

[3] For instance, the authors explicitly affirm the importance of both: after introducing the conditions of autoregressive and moving average parameters separately, Box and Jenkins note that a "mixed process of considerable practical importance" is the ARMA(1, 1) (Box and Jenkins, 1976, 76).

identification, fitting, and diagnostic checking to arrive at the appropriate model (Box and Jenkins, 1976, x). This iterative process was challenging. Users had to estimate starting values based on a combination of sample values, recommended Appendix starting values, and the ARIMA differenced equation. This complicated process was "particularly well suited for use on a electronic computer" and sensitive to the choice of starting values in the world of small n (Box and Jenkins, 1976, 212, 213). At the time of the introduction of the modeling strategy, many analysts found themselves without one (the computer) and with the other (short n), especially in the social-scientific world. Even after estimation, the likelihood function often suggested "two sets of value of the parameters which might explain the data," which they called the "estimation situation" (Box and Jenkins, 1976, 225).

Given these practical constraints, the Box-Jenkins methodology might have been dismissed early on as too impractical or obtuse by applied social scientists. A casual reading of the approach, though, offered an apparent way out: Box and Jenkins make it clear that the former half of "ARMA" is much easier to estimate. For autoregressive parameters, except for the effect of starting values, the series is "linear in the ϕ's"; this linearity lends itself to well known and relatively simple estimators like ordinary least squares to estimate the values. For moving average parameters, the series are "always nonlinear functions of the parameters" (Box and Jenkins, 1976, 232).[4] This explicit invitation to simplicity is at odds with the involved moving average estimation process.[5] Even though, at the time, general nonlinear estimation routines were "becoming generally available" (Box and Jenkins, 1976,

[4]Moreover, for autoregressive (not mixed) processes, Box and Jenkins devote Appendix 7.5 to demonstrating how autoregressive estimates could arise from linear least squares or two maximum likelihood methods; they find that "the estimates given by the various approximations will be small ... we normally use the least squares estimates" (Box and Jenkins, 1976, 279).

[5]None of this is to suggest that Box and Jenkins ever advocated that users focus on the autoregressive component instead of the moving average. Quite the opposite. The authors take great care to note that the autocorrelations themselves will still *depend* on the MA process, where the order of the MA process determines the respective autocorrelations. To Box and Jenkins, "for the ARMA(p, q) process, there will be q autocorrelations $\rho_q, \rho_{q-1}, \ldots \rho_1$ whose values depend directly on the choice of the q moving average parameters" (Box and Jenkins, 1976, 75). So the advice is *not* that we can remove MA parameters and achieve a "good enough" specification; rather, the interplay is obvious.

233), this does not speak at all to their availability to social scientists.

Other specialized language might have further clouded the problem. When speaking about the *autocorrelation functions* of the AR and MA process, Box and Jenkins note a "duality between the AR(1) and MA(1) processes" (Box and Jenkins, 1976, 70). Later, when discussing the generality of the representation of the ARMA model, Box and Jenkins note the "multiplicity" of the ARMA model, where there are "multiple solutions for moving average parameters obtained by equating moments" (Box and Jenkins, 1976, 198). The nature of this multiplicity is immediately clarified—there is only one *stationary invertible* solution for the ARMA parameters (Box and Jenkins, 1976, 197)—but a casual reader might mistake this section, in conjunction with others, as advice for the multiplicity of alternatives for the *model in general.* Especially given the iterative approach to model building advocated overall, users, especially limited by computing resources, might be tempted to jettison the (complicated) moving average component for the (simpler) autoregressive component.

This invitation is even louder in the face of potential redundancy or duplicity in the ARMA specification. Box and Jenkins warn that, if the analyst fits an ARMA model where the autoregressive and moving average components are potentially motivated by a common factor,[6] the estimated model could result in "extreme instability in the parameter estimates" due to the near-cancellation of the factors (Box and Jenkins, 1976, 248). The authors follow this situation of *extreme near cancellation* with two clarifications. First, speaking of the AR and MA components, they note that "a change of parameter value on the left can be nearly compensated by a suitable change on the right" (Box and Jenkins, 1976, 248). Of course, what they mean is that suitable *combinations* of the AR and MA estimates might yield identically good solutions to the likelihood function, but it could be casually interpreted that the left- and right-hand sides of the equation are *compensatory*, so the analyst only need include one. Second, noting the identical emerging interpretations from this extreme

[6]The authors give the example of $(1 - 1.3B0.4B^2)\tilde{w}_t = (1 - 0.5B)a_t$ (Box and Jenkins, 1976, 248).

cancellation situation, Box and Jenkins recommend "*parsimony* in parameterization," and explicitly recommend "the fitting of the AR(1) process, which would normally be entirely adequate" (Box and Jenkins, 1976, 249, emphasis in original).[7] This, of course, flies in the face of the unique role of the MA process in generating errors, but an analyst looking to dump the moving average in favor of the autoregressive component might readily heed the call. In the face of limited computing resources, a short t, potential model multiplicity, and now saw that these equations have multiple reasonable solutions, earlier generations of scholars might reasonably respond by focusing on solutions which retain only the AR properties of the series.

In later revised editions, Box, Jenkins, and Reinsel would clarify that these estimation routines "are now generally available" (Box, Jenkins and Reinsel, 1994, 250). And in discussing the estimation strategies themselves, the authors shift to note "it is scarcely worthwhile" to separate out autoregressive processes to use the ordinary least squares estimator (Box, Jenkins and Reinsel, 1994, 302). Instead, the whole ARMA model should be deferred to a maximum likelihood estimation program, which, of course, is exactly what we recommend here. But a few words in an Appendix in 1994 is a small force in the face of 20 years of autoregressive modelling practices.[8]

Clearly, both the AR and MA components played critical roles in the original Box-Jenkins formulation. For practitioners who "grew up" with the Box-Jenkins transfer function methodology, we can explicitly state that the method offers obvious leverage in how it handles modelling the dynamics of dependent variables, both with observed covariates and the error term. But for a combination of practical and historical reasons, the theoretical leverage of the approach has been lost, leading to a broad inertia behind estimating autoregressive

[7]This is perhaps exacerbated by the statistical observation that the MA components of a mixed series do not affect the stationarity of the autoregressive process, which is determined whether the the equation $\phi(B) = 0$ has roots outside the unit circle (Box and Jenkins, 1976, 74).

[8]The authors are aware of the double meaning of the forgoing sentence and are quite pleased with themselves. Readers are permitted a groan.

models at the expense of moving averages.

5 Moving Averages and Hendry

The dominant approach to modelling dynamic processes in political science is Hendry (1995) and the general-to-specific methodology. Hendry advocated for identifying the local data-generating process of a dependent variable by encompassing it in a general, unrestricted model and then paring down said model to recover the local data-generating process.[9] Political scientists normally pare down the model themselves, while economists rely on automated procedures. Extensive simulations have demonstrated that this method recovers the local data-generating process nearly as often as if the scholar, using a priori knowledge, had estimated the local data-generating process directly (Hoover and Perez, 1999; Hendry and Krolzig, 2004). Such findings strongly recommend this process for identifying the dynamic structure of a time series.

The general model Hendry recommends for capturing dynamics in a time series is an autoregressive distributed lag model. Enough lags of the dependent and independent variables need to be included in order to completely model the dynamics in the data, as confirmed by a post-estimation test such as the Breusch-Godfrey. As has long been known in economics, this model accommodates autoregressive errors as well as lags of the dependent and independent variables. Moving averages are excluded from these models, though it has long been acknowledged that a generalization of autoregressive distributed lag models would include moving averages (Hendry, Pagan and Sargan, 1984). Thus while the dynamic regression models estimated using the general-to-specific approach are somewhat flexible, they do not achieve the same level of flexibility as transfer function models in the Box-Jenkins approach.

(Hendry, 1995) explicitly justifies the exclusion of moving averages from this framework

[9]For a complete review of the methodology, see Hendry (1995).

in terms of identification; that a general-to-specific framework cannot accommodate moving averages because common-factor restrictions will result in multiple parameter combinations that perfectly capture the data. As the above analysis indicates, however, there is a distinct difference between the identification issues with autoregressive errors and moving averages. A dynamic model with autoregressive errors can be perfectly reexpressed using a finite set of lags of the dependent and independent variables. A dynamic model with moving averages, however, can only be perfectly reexpressed using an infinite set of lags of the dependent and independent variables, a model that is infeasible with a finite set of data. Thus, there is no risk of creating identification issues when including moving averages in a general-to-specific modelling strategy.

Implicitly, Hendry justifies the exclusion of moving averages through the approximation of moving averages using autoregressive errors. Hendry and Trivedi (1972) tested the effect of approximating moving average errors with autoregressive errors and vice versa. They found that models with autoregressive errors of the same order as the moving average they are meant to represent can lead to remarkably accurate models, with RMSE comparable to a model of that models the true data-generating process using moving averages. Approximating moving averages also reduces the bias associated from not modelling moving averages whatsoever, though the bias reduction is not close to the true model. With forecasting as a predominant goal, Hendry (1995) concludes that moving averages should just be approximated with autoregressive errors of the same order.

While such a conclusion may be appropriate in economics, it is inadequate in political science given the importance of inference rather than prediction. The bias that remains from approximating moving averages may create problems for inference of both short-run and long-run relationships between the dependent and independent variables, which are much more important to scholars than measures of total model fit. Given that the estimation is now straightforward, dynamic regression models with moving averages are much more compelling to estimate rather than relying on a potentially large number of lags of the dependent and independent variables to approximate moving averages.

Even if ADL models could calculate quantities of interest with a reasonable degree of accuracy, doing so would be inefficient. Approximating moving average process with lags of the dependent and independent variables requires a larger number of parameters than the moving average process it is meant to approximate. This unnecessarily wastes degrees of freedom, resulting in larger standard errors and higher chances of false negative results in the data. This problem will further be exacerbated if the number of independent variables in the analysis is large. Given the focus on inference for political science, the inclusion of moving averages in dynamic regression models is optimal.[10]

6 Simulations

We revisit the effect of omitting moving averages through simulations. Specifically, we simulate 1,000 iterations of a dynamic data-generating process represented by equations 1, 2, and 6; in effect, the data-generating process is a partial adjustment model with a moving average error term.[11] α and β are both equal to 0.5, while ρ is equal to 0.95.[12] ε_{1t} and ε_{2t} are white noise error processes drawn from a standard normal distribution. Both θ and the number of time points, t, are manipulated to varying degree in order to best demonstrate the effect of omitting a moving average. All series have a burnin period of 50.

For each dataset, a variety of models are estimated. A partial adjustment model is estimated in order to demonstrate the basic criticisms of Achen (2000). An ADL(2,1) model is estimated to demonstrate the utility of approximating moving averages with autoregressive

[10] One might be tempted to extend this logic to the direct estimation of autoregressive errors rather than lags of the dependent and independent variables as suggested by Wilkins, but this would lead to identification problems.

[11] As opposed to an autoregressive error term, as in the Achen papers.

[12] These values are chosen in an attempt to mirror Wilkin's parameter values. We have investigated changes to α and β and found the results are substantively similar.

errors, as advocated for by Hendry (1995). An ADL(5,5) model is estimated to demonstrate how an applied scholar might attempt to soak up all possible autocorrelation in the residuals using the general-to-specific modelling approach. Each of these models are then estimated again, this time including a first-order moving average parameter to demonstrate the utility of their inclusion. Finally, a static model is also included for comparison.

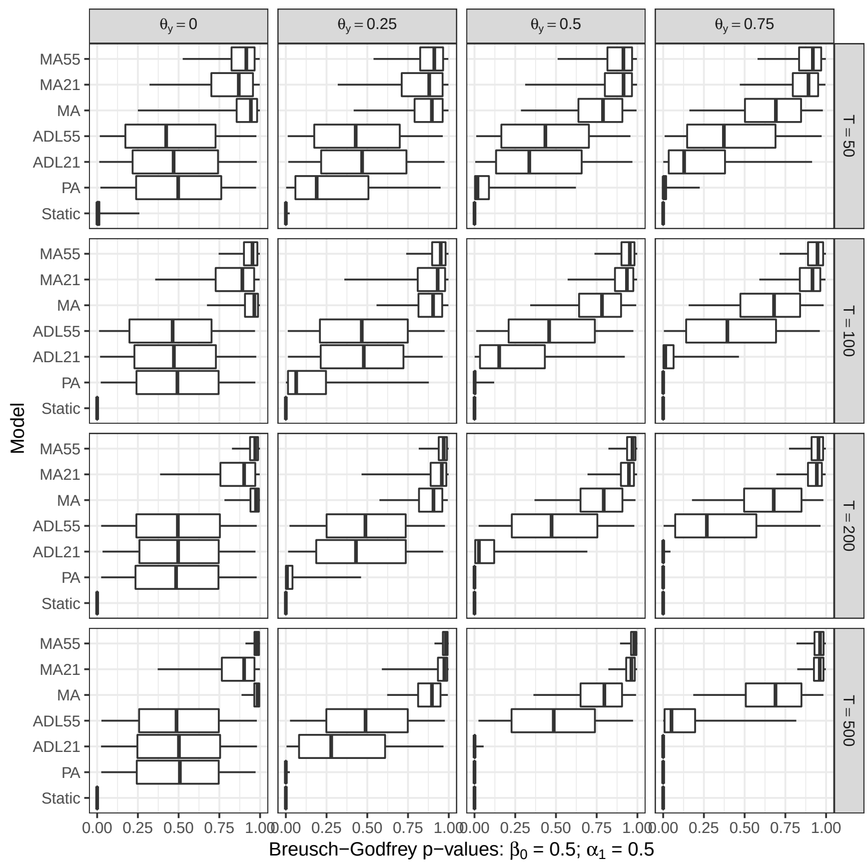

Figure 1: Breusch-Godfrey Test Results ($\alpha = 0.5$).

Figure 1 presents the distribution of p-values of a Breusch-Godfrey test for serial autocorrelation in the residuals for each of the estimated models. To begin, we examine the leftmost column in which θ equals zero, or there is no moving average. For every sample size, the static model has consistently small p-values. Thus, the Breusch-Godfrey test consistently rejects the null hypothesis of no serial autocorrelation in the disturbance term. In contrast, all of the other models have consistently higher p-values that fail to reject this same null hypothesis. Both of these results are expected, but they are pointed out to anchor the discussion of the rest of the figure.

There are two trends that become apparent when looking at the graphic. As the strength of the moving average θ increases, the tests more consistently reject the null hypothesis of no serial autocorrelation in the disturbance term for the models without moving averages. In most scenarios, the PA model will reject the null hypothesis of a lack of autocorrelation. But the ADL(2,1) model will often fail to reject the null hypothesis with either a small sample or a small θ, with the conclusion being that the setup adequately models all autocorrelation in the data. Only the ADL(5,5) consistently fails to reject the null hypothesis that the residuals are white noise error terms.

These results indicate that scholars following currently accepted guidelines for modelling dynamics without moving averages will either rely on approximating a moving average with an autoregressive error setup or with a large number of lags. Of course, residual autocorrelation will remain regardless of which model is chosen, as a finite number of lags cannot full capture a moving average process; only the models with moving averages ensure there is no residual autocorrelation remaining. But it's worth pointing out for the purposes of examining which models we should focus on moving forward.

Figure 2 shows the simulation results for estimating β_0. The red line represents the true value of β in the simulation. Unsurprisingly, the static model, which has no dynamics, returns biased estimates of β. The PA model also has notable bias as both θ and t increase. The remaining models are all largely unbiased, with the median estimate almost exactly lining up with the true parameter. The sampling distributions for these estimators are quite different, however, with variance in the parameter estimates declining as the number of estimated parameters decline. The correct model, of course, has the fewest parameters and thus the smallest sampling distribution. This is unsurprising, as the moving average model

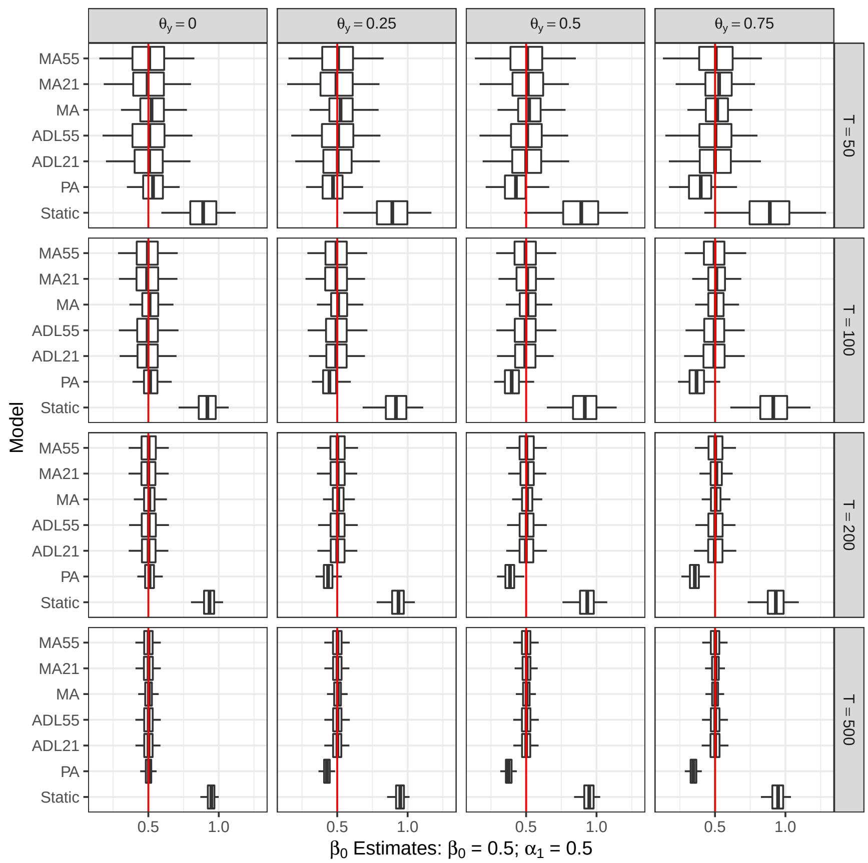

Figure 2: β Estimates ($\alpha = 0.5$; True $\beta = 0.5$).

contains fewer parameters than the other models of the data.

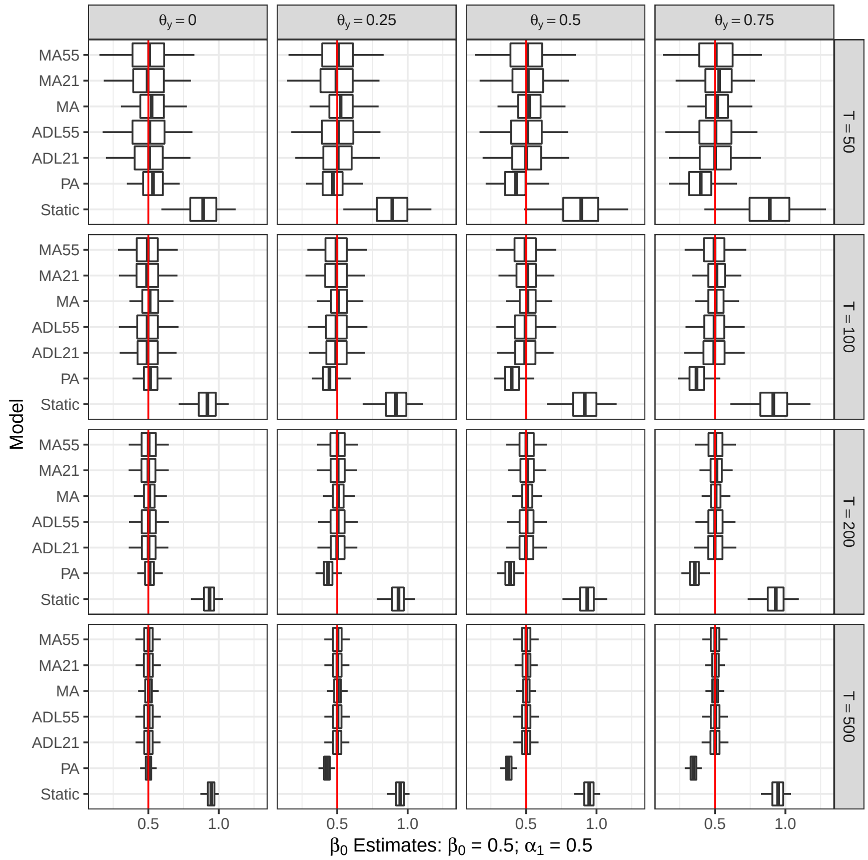

Figure 3: β Standard Errors ($\alpha = 0.5$; True $\beta = 0.5$).

Figure 3 shows the simulation results for the standard errors of β_0. Again, we see that the correct model has a much smaller standard error than the larger ADL models. The PA model has the smallest standard error, which is unsurprising because it has even fewer parameters than the moving average model. This model is biased, however, so the smaller standard error does not indicate it should be preferred. These combined results indicate that for scholars concerned about inference of the short-run marginal effect of x on y, moving average models are preferable to lag approximation because moving average models will have a smaller sampling distribution.

Figures 4 and 5 show the distribution of the long-run multiplier estimates and their

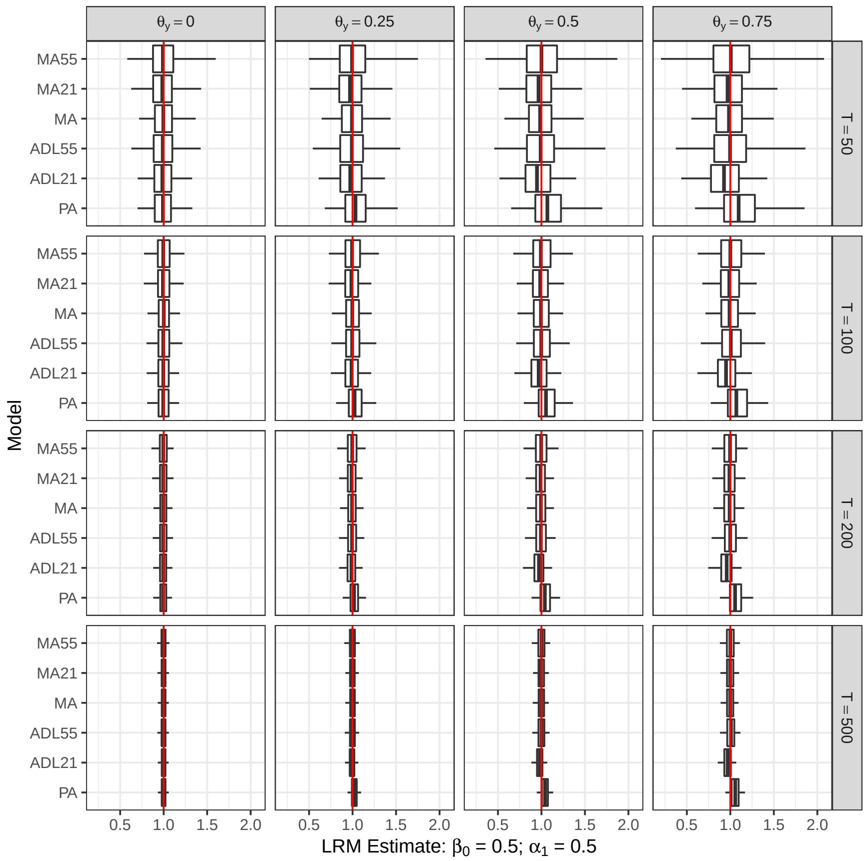

Figure 4: Long-run Multiplier Estimates ($\alpha = 0.5$).

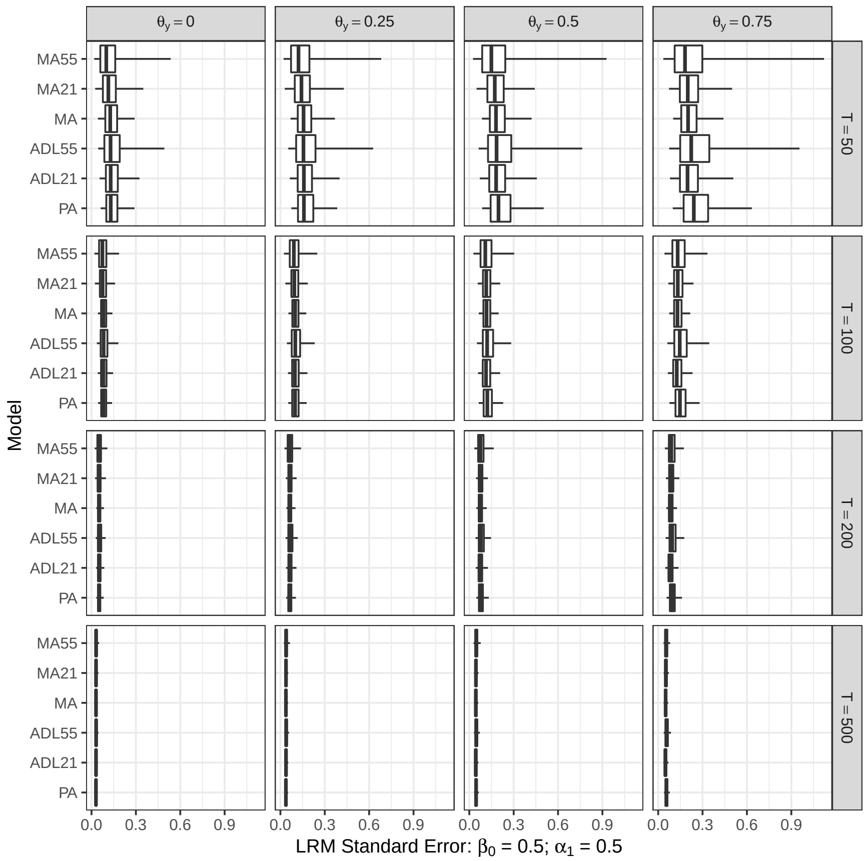

Figure 5: Long-run Multiplier Standard Errors ($\alpha = 0.5$).

standard errors. The results largely reflect those of the estimates for β. The ADL(2,1) and ADL(5,5) estimates are largely unbiased, as are the estimates from the moving average models. However, the correct model only has a slightly smaller sampling distribution. The advantage of the moving average model is more limited here than in the previous case: it is only an apparent advantage in small samples and only has a distinct advantage over the ADL(5,5). Still, a model with a moving average is preferred to lag approximation.

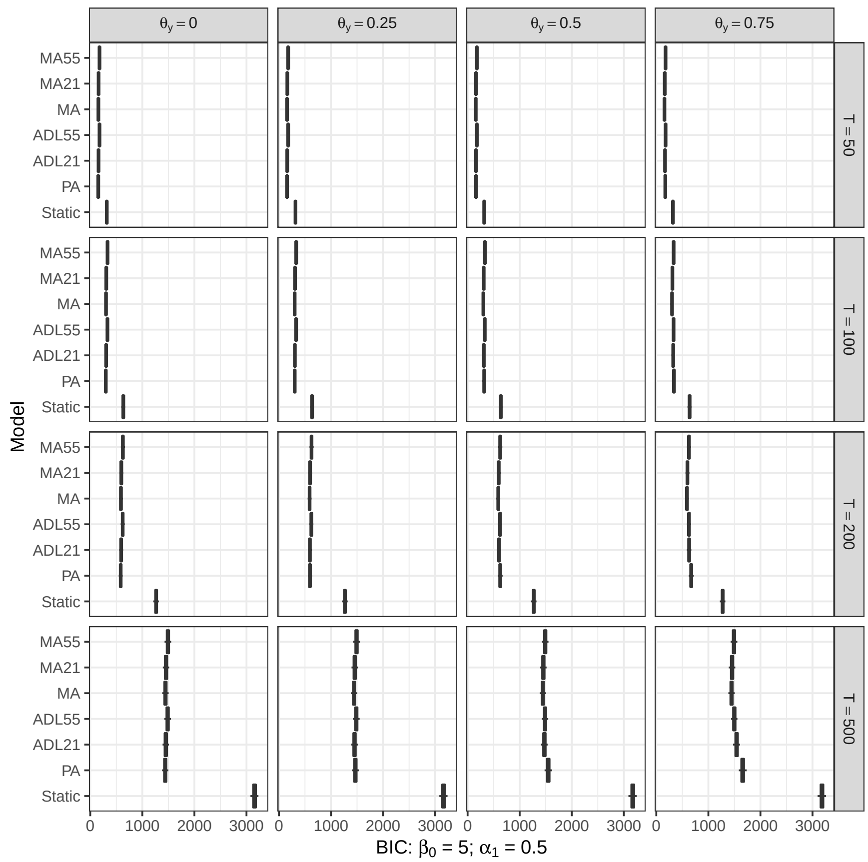

Figure 6: BIC ($\alpha = 0.5$).

Figure 6 indicates the BIC of the estimated models. Overall, the only discernible pattern from the BIC is that the static models are wholly inadequate. Beyond that, it is difficult to tell whether any other model has a substantial benefit in terms of overall explanation of a sample of data. This is consistent with Hendry and Trivedi (1972)'s finding on the matter that approximating moving averages with autoregressive errors (or accomplishing the same with lags of observed variables) leads to an overall model fit that is largely equivalent to the true model.

7 Incorporating Moving Averages Into Dynamic Regression Models

The Monte Carlo results demonstrate that moving averages should be included in dynamic regression models when a scholar is concerned about inference. Of course, this broad recommendation leaves much room to interpretation in how moving averages are incorporated into dynamic regression models. We begin with advice on how to estimate and interpret dynamic regression models with moving averages, later moving to describing how such models fit into varying approaches to modelling time series data.Our hope is that such advice will make moving averages accessible to future scholars.

Most standard software packages allow scholars to estimate ARIMA models as part of their base offering, including R and Stata. Such programs allow scholars to include independent variables as part of their model, in what are often described as ARIMAX models. These functions allow scholars to estimate a dynamic regression model with moving averages, which is what we recommend. But scholars should be careful to properly understand the functions they employ. When estimating an ARIMA model in R or Stata using base commands while including independent variables, the specified ARMA components of the model will be calculated for the error term rather than the dependent variable. Thus, scholars should recognize that the AR component that is available for scholars to specify refer to autoregressive errors, not lagged dependent variables. To prevent identification issues resulting from autoregressive errors, these should be set to zero. Lags of the dependent variables must still be included manually among the independent variables.

Interpreting models with moving averages is straightforward, as one does not have to change how one interprets dynamic regression models in order to accommodate moving averages. Short-run and long-run marginal effects retain their same formula and interpretation. Moving average terms themselves should not be interpreted, as it is still part of the error term and interpreting the the error term is a fool's errand. Moving averages can largely be regarded as a nuisance parameter that can be estimated and ignored when statistically significant and omitted if not.

In terms of modelling approaches to time series data, we begin with a description of how to incorporate moving averages into the general-to-specific modelling strategy. When using autodistributed lag models in the general-to-specific approach, the general recommendation is that scholars craft a general model that includes the same number of lags of the dependent and independent variables, testing to determine whether this adequately captures the dynamics in the data. We extend this recommendation to also include moving averages, with the number of moving averages equal to the number of lags of the observed variables. In most political science applications, this will mean a single moving average. However, many more moving averages are certainly possible, especially with seasonal data. [13]

Users of automated methods for general-to-specific modelling may object to this approach on the grounds that it increases computation time. After all, a maximum likelihood model takes longer to estimate an ordinary least squares model, and the number of models an automated method may calculate can enter into the hundreds or thousands. These objections are practical, not theoretical; they are only relevant if available computing power is insufficient. But while these objections might have been relevant when automated methods were first introduced, computing power has greatly advanced. Parallel processing is commonplace at universities, increasing the speed of repetitive computations tremendously. Many scholars

[13] It is possible that including many higher-order moving averages may increase computation time and potentially preventing convergence in traditional maximum likelihood estimators. In this event, Bayesian MCMC estimation could easily estimate a dynamic regression model with moving averages.

can perform such processing through their personal computers. Thus, this objection seems shallow for most researchers.

Some scholars object to general-to-specific methods in favor of theoretically driven specifications of their models. It is somewhat difficult to recommend how to choose moving averages from theory alone, as the error term includes all possible variables not incorporated in the model. It is known, however, that measurement error in a lagged variable, whether dependent or independent, will result in a moving average process in the error term. Thus, if scholars are using theory to justify the inclusion of particular lags, they should also include moving averages of the same order(s) to account for the possibility of measurement error.

8 Conclusion

Moving averages can cause the same problems for dynamic regression models as autoregressive errors, which is bias in quantities of interest when it is unmodelled. But unlike with autoregressive errors, the autocorrelation from moving averages cannot be perfectly captured by lags of the dependent and independent variables. Even when moving averages are approximated via lags, estimates of quantities of interest will be inefficient. For scholars concerned with inference, then, moving averages need to be part of their dynamic regression toolkit.

The discussion of moving averages in the time series context begs the question of moving averages in the spatial regression context. It has long been known that the two estimation procedures are quite similar in both vernacular and technical details (Fingleton, 2009). Indeed, the fact that a spatial error model can be reexpressed as a spatial Durbin model is a result of the fact that the only spatially autocorrelation currently modelled in error terms is spatial autoregression; spatial moving averages are rarely considered (Anselin, 2003). These results suggest that spatial moving averages need to be included in spatial regression models as well, though such a claim would need to be verified upon further investigation.

Indeed, both temporal and spatial moving averages also deserve attention in the time series cross-sectional context. There is no reason to believe that either process ceases to be relevant once data is analyzed across both time and space. But such processes are rarely modelled, likely due to the complexity of capturing moving averages alongside autoregressive processes and unobserved effects. More research needs to be done on the estimation and interpretation of such models.

References

Achen, Christopher H. 2000. "Why lagged dependent variables can suppress the explanatory power of other independent variables." *Ann Arbor* 1001(2000):48106–1248.

Anselin, Luc. 2003. "Spatial externalities, spatial multipliers, and spatial econometrics." *International Regional Science Review* 26(2):153–166.

Beck, Nathaniel. 1985. "Estimating dynamic models is not merely a matter of technique." *Political Methodology* 11:71–89.

Box, George E. P. and Gwilym M. Jenkins. 1976. *Time Seires Analysis: Forecasting and Control*. John Wiley and Sons.

Box, George E. P., Gwilym M. Jenkins and Gregory C. Reinsel. 1994. *Time Seires Analysis: Forecasting and Control*. John Wiley and Sons.

Fingleton, Bernard. 2009. "Spatial Autoregression." *Geographic Analysis* 41(4):385–391.

Hendry, David E. 1995. *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, David F, Adrian Pagan and J. Denis Sargan. 1984. "Dynamic Specification." *Handbook of Econometrics* 2:1023–1100.

Hendry, David F. and Hans-Martin Krolzig. 2004. "Automatic model selection: a new instrument for social science." *Electoral Studies* 23(3):525–544.

Hendry, David F. and P. K. Trivedi. 1972. "Maximum likelihood estimation of difference equations with moving average errors: A simulation study." *The Review of Economic Studies* 39(2):117–145.

Hoover, Kevin D. and Stephen J. Perez. 1999. "Data mining reconsidered: encompassing and the general-to-specific approach to specification search." *Econometrics Journal* 2:167–191.

Keele, Luke and Nathan J Kelly. 2006. "Dynamic models for dynamic theories: The ins and outs of lagged dependent variables." *Political Analysis* 14(2):186–205.

Philips, Andrew Q. 2018. "Have Your Cake and Eat It Too? Cointegration and Dynamic Inference from Autoregressive Distributed Lag Models." *American Journal of Political Science* 62(1):230–244.

Shumway, Robert H. and David E. Stoffer. 2017. *Time Series Analysis and Its Applications*. Springer.

Slutsky, Eugen. 1937. "The Summation of Random Causes as the Source of Cyclic Processes." *Econometrica* 5(2):105–146.

Wilkins, Arjun S. 2018. "To Lag or Not to Lag?: Re-Evaluating the Use of Lagged Dependent Variables in Regression Analysis." *Political Science Research and Methods* 6(2):393–411.