Chapter 5: Regression Models for Stationary Time Series

February 18, 2022

Contents

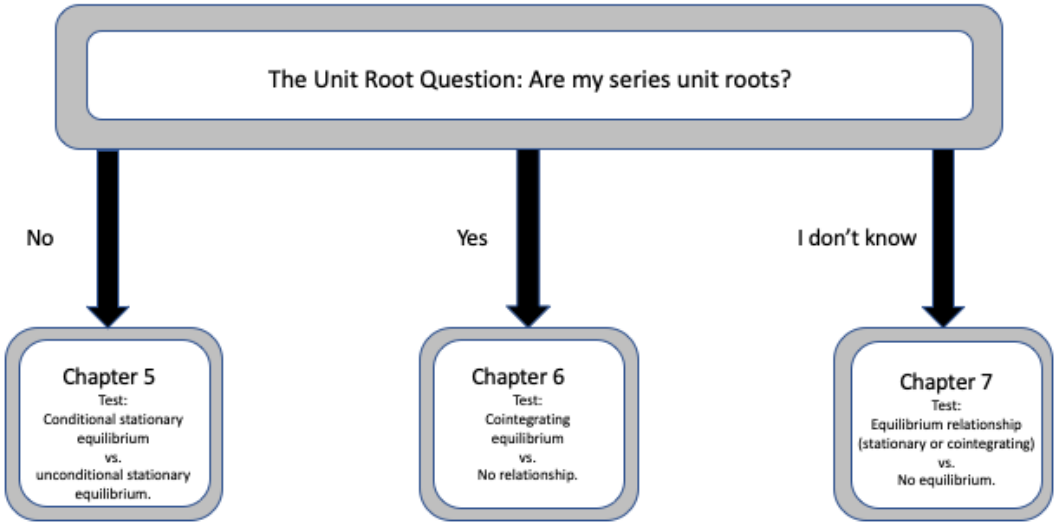
Introduction to Part II

Armed with knowledge about their univariate properties and the type of exogeneity that may reasonably be assumed, you are ready to conduct multiple regression analysis with your time series data. Part II tackles estimation and inference with dynamic regression models assuming weak exogeneity.

As we delve into regression models, it is useful to think about the concept of an equilibrium in time series analysis. Banerjee et al. (1993, p.2) explain that a system is in an equilibrium state when there is no tendency to change over time. Since we are interested in dynamics, we do not focus on whether series are at an equilibrium point. Rather, we study whether series are moving towards equilibrium and what factors account for that movement.

As we explained in *Beyond the Unit Root Question* (**?**), equilibria can take various forms and not all time series have one. The term *unconditional stationary equilibrium* refers to the idea where a stationary variable returns to its mean and is not determined by exogenous regressors (**?**, p.279). Such a series likely reacts to factors that are not yet identified as independent variables. A *conditional stationary equilibrium* occurs when the dependent variable is stationary and also is a function of exogenous regressors in the model.

Figure 1: A Map of Part II



The map of Part II in Figure 1 shows the layout of the next three chapters and the types of hypotheses we will cover in each one. In Chapter 5 we focus on the analysis of stationary data. We are interested in testing for conditional stationary equilibria against the possibility of unconditional stationary equilibria. To do this, we introduce a typology of regression models appropriate for stationary time series. Many people think of the implements of time series analysis as econometric shortcuts. For example, lagged dependent variables are used to "control for serial correlation," or the regressors are lagged one period to "resolve the endogeneity problem." But dynamic modeling is more complicated than that. When lags are included in a model, the easy correspondence between the effects of the regressors and estimated slope coefficients is broken. Lags of the regressand mean that shocks to the regressors propagate into the regressand, evolving over time. Different combinations of these lags imply different things about the nature of the underlying relationship. Chapter 5 explains how one fits regression models to stationary data, outlines the new quantities of interest that become relevant when one moves into the realm of dynamic regression modeling, and offers strategies for developing a dynamically complete specification of a hypothesized relationship.

In contrast to a stationary series, a single unit root series does not have a mean to which it reliably returns – rather, the mean changes over time and the series' variance grows. A unit root series cannot have its own equilibrium but it can have one relative to another unit root series. That is, the linear combination of more than one unit root series can be stationary. This third type of equilibrium is a *cointegrating equilibrium* and can occur among series that all contain unit roots. For analyses where the dependent variable is a unit root and we can assume weak exogeneity of the regressors, Chapter 6 introduces appropriate single-equation methods to search for a cointegrating equilibrium. We cover the concept of cointegration and explain how to test for its presence, how to estimate cointegrated regression models, and how to draw inferences from the resulting models. We'll also cover what to do when there is no evidence of cointegration.

As we saw in Part I, it is often not possible to be certain about whether a series contains a unit root. The tests we rely on to make such choices can produce inconsistent results and lead to uneasy conclusions about stationarity and the unit root question. That uncertainty is usually forgotten as one chooses a regression model. Until recently, analysts were forced to choose between the tools offered in Chapter 5, assuming stationary time series, and those in Chapter 6 that assume unit root time series. But, without certainty on the unit root question, you might not know if you are looking for a conditional stationary equilibrium or a cointegrating equilibrium. Thus, in Chapter 7 we present two bounds approaches that provide a more transparent framework for inference when you are uncertain about the classification of at least some of your data.

Another important concept for the time series regression models in Part II is equation balance. **?** provides an intuitive explanation of the concept. A "minimum condition" for a time series regression model to be "satisfactory is that the explanatory variables should be capable of explaining" all the dominant features of the explained variable (Granger 1991, 12-13). If y_t contains a unit root, there must be at least one regressand that explains the non-stationary behavior of y_t for the equation to be balanced. If y_t contains a deterministic trend, this deterministic trend must be explained by one of the independent variables or a trend must be added as a regressor for the equation to be balanced. All of the dominant features of the regressand must be explained by the set of regressors and all the dominant features of the regressors must be reflected in the regressand. If these conditions are not met, the model cannot be correctly specified.

With the concept of equation balance at the front of your mind, you will apply the same general approach to model building in each of the next three chapters. Theory informs the variables that are included in the model. What is the dependent variable, what are the independent variables of interest, and what are the variables that need to be included as controls? The data will guide you toward an appropriate dynamic specification. You begin with a plausible general model – a model that is balanced and contains a sufficient number of lags to encompass all they dynamics in the data – and you test restrictions on the general model until you arrive at a specification that is parsimonious and dynamically complete.

Diagnostic testing plays an important role in this iterative specification procedure but so does good judgement. The tools we discuss in Part II should not be treated as flawless instruments that provide definitive answers. Though we have done our best to provide road maps to help you navigate the choices you have to make when applying these tools, the way forward will not always be clear. When these situations arise; you should be as transparent as possible about the choices you made and why you made them and you should consider using multiple models and specifications to evaluate the robustness of your findings.

The approaches outlined in Chapters 5, 6, and 7 represent what we consider to be best practices for applied time series analysis for stationary and non-stationary time series data. We provide examples that will demonstrate how these tools are used in applied settings. As you navigate these chapters it will become clear that your theory, the information that can be derived from your data, and your common sense are all essential.

Chapter 5

In Part I, we discussed the preliminary steps of applied time series analysis. If your work to this point suggests that (at least) your outcome variable of interest is stationary (or trend stationary) and that the explanatory variables

of interest are at least weakly exogenous then this chapter will provide the tools you need to build single-equation regression models, draw inferences, and interpret your findings.

The model building process requires that you begin with a plausible general model that encompasses all the dynamic features that may be present in the data. You use theory and previous research to determine which variables belong in the model. Because the dynamics you are likely to observe in your time series vary based on your sampling window and sampling interval, the data are left to determine the dynamic structure of the relationship. Using the information you have gleaned from preliminary analysis and a series of diagnostic tests, you will iteratively test restrictions on the general model until you arrive at a specification that is both parsimonious – there are no unnecessary lags in the model that might reduce the power of your hypothesis tests – and dynamically complete – the final model meets all the assumptions required to support hypothesis testing.

We begin the chapter by discussing the classical linear regression assumptions with cross-sectional data in Section 1. These assumptions imply that OLS estimators are BLUE (best, linear, unbiased estimators). We then introduce the weaker assumptions that apply in the time series case when lagged values of the regressand appear on the righthand side of the model in Section 2. When these assumptions are met, OLS is not longer BLUE but is consistent. Given this foundation, we describe the model-building process for stationary time series data in Section 3. First, we specify a general model in Section 4. We introduce two general models that have been widely used in the social sciences to model a stationary outcome variable. Both models assume the same underlying dynamic structure but estimate different quantities directly. We begin with the autoregressive distributed lag model (ADL) in Section 4.1 and then discuss the generalized error correction model (GECM) in Section 4.2. In step two, we test for violations of the time series regression model. Section 5 we introduce tests designed to help you evaluate whether the general model satisfies the time series regression assumptions. Then, in Section 6, we discuss five models that impose various restrictions on the parameters of the ADL and GECM – and thus the dynamic structure underlying the relationships in the data – and describe how to test the validity of the restrictions. From here, we move to interpreting relationships in Section 7. We explain the quantities of interest for hypothesis testing and how to draw rich interpretations from your results. In Section 8, we present examples. We end the Chapter with XXX in Section 9.

1 Multiple Regression With Cross-Sectional Data

In a cross-sectional context, ordinary least squares (OLS) regression analysis proceeds in three parts. You begin by specifying a theoretical model. Consider the following three-variable regression model.

$$y_i = \alpha_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \mu_i. \tag{1}$$

The dependent variable, or regressand (y_i) is specified as a linear function of two independent variables, or regressors, x_{1i} and x_{2i}. The i subscripts denote variation over cross-sectional units.

The model parameters α_0, β_1, and β_2 govern the relationships between the regressand and the regressors. Unit changes in x_{1i} or x_{2i} are hypothesized to produce β_1 and β_2 changes in y_i, respectively, and α_0 represents the mean of y_i when $x_{1i} = x_{2i} = 0$. The population error term μ_i is included to highlight that there is variation in the regressand that cannot be completely explained by the hypothesized model. We tend to think of this hypothesized population regression function (PRF) as theoretical because the population parameters α_0, β_1, and β_2 cannot be calculated unless the analyst has access to the full universe of cases from which the variables y_i, x_{1i}, and x_{2i} are drawn. In most applied settings, we only have a subset of this universe, our sample.

We use the sample regression function (SRF) to make inferences about the PRF. The SRF for the hypothesized relationships is shown in Equation 2.

$$y_i = \hat{\alpha}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\mu}_i. \tag{2}$$

The equation for the SRF is similar to the PRF. The only differences between Equations 1 and 2 are the *hats* ($\hat{\square}$) that appear above the parameters; $\hat{\alpha}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$; and the sample error term (or residual) $\hat{\mu}_i$. The quantities $\hat{\alpha}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ are sample regression coefficients. These coefficients are estimated by minimizing the sum of the squared differences between the observed values of the regressand (y_i) and the values of the regressand predicted by the model ($\hat{y}_i$), the

sum of the squared sample residuals ($\sum \hat{\mu}_i^2 = \sum (\hat{y}_i - y_i)^2$). This criterion explains the *ordinary least squares* (OLS) moniker.

Because the sample regression coefficients are estimated from a sample, the sample regression coefficients are subject to variability. The variances ($\hat{\sigma}^2_{\hat{\beta}_1}$, $\hat{\sigma}^2_{\hat{\beta}_2}$, $\hat{\sigma}^2_{\hat{\alpha}_0}$) and standard errors ($\hat{\sigma}_{\hat{\beta}_1}$, $\hat{\sigma}_{\hat{\beta}_2}$, $\hat{\sigma}_{\hat{\alpha}_0}$) for these coefficients can be estimated from the sample to quantify this variability. The variances and standard errors of the coefficients are a function of the variance $\hat{\sigma}^2_{\hat{\mu}_i}$ and standard error $\hat{\sigma}_{\hat{\mu}_i}$ of the regression, quantities that summarize the variability of the estimates.

Inference typically proceeds in one of two ways. If you have formulated hypotheses about the individual population parameters, the coefficients and the standard errors for the coefficients can be used to produce a series of test-statistics:

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}}, \quad t_{\hat{\beta}_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\hat{\beta}_2}}, \text{ and } t_{\hat{\alpha}_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\hat{\alpha}_0}}. \tag{3}$$

If you are testing hypotheses that postulate null relationships between the regressors and the regressand, and/or a hypothesis that the y_t-intercept (α_0) is 0, the test statistics in Equation 3 reduce to the ratios of the estimated coefficients to their estimated standard errors.

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}}, \quad t_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{\hat{\sigma}_{\hat{\beta}_2}}, \text{ and } t_{\hat{\alpha}_0} = \frac{\hat{\alpha}_0}{\hat{\sigma}_{\hat{\alpha}_0}}. \tag{4}$$

These test statistics are compared to appropriate t-distributions that are selected based on model degrees of freedom, the difference between the number of observations in the sample less the number of parameters estimated ($N - k$). You infer that the parameter is reliably different from zero if the t-statistic exceeds the associated critical value. You can also use a number of tests to evaluate joint hypotheses on the coefficients including F-statistics, Lagrange Multiplier statistics, and Likelihood-Ratio test statistics.[1]

The Gauss-Markov assumptions establish the optimal conditions for the OLS estimators. The enumeration of these basic assumptions varies across texts. Regardless of how the assumptions are outlined, OLS is said to be the best linear unbiased estimator (BLUE) when the Gauss-Markov assumptions are met. For ease of exposition we will define $\mathbf{x_i}$ as a set of regressors x_{ji}, where the subscript ji denotes one of $j = 1, ..., k$ explanatory variables for observation i. Given this notation, the critical assumptions for cross-sectional regression models are as follows.

CS1. Linearity: As the name implies, the model is assumed to be linear in parameters. This means the joint data generating process for $\{(\mathbf{x_i}, y_i) : i = 1, 2, \ldots, N\}$ follows the linear model given by

$$y_i = \mathbf{x_i}\boldsymbol{\beta} + \mu_i \tag{5}$$

where $\{\mu_i : i = 1, 2, \ldots, N\}$ is the sequence of disturbances and N is the number of observations. A model that includes terms like $\beta_1 x_1 x_2$ or $\beta_2 x_2^2$ is still linear in parameters. A model that includes terms like $\beta_1 \beta_2 x_1$ or $\beta_2^2 x_2$ is not.

CS2. Zero Conditional Mean: $E(\mu_i|\mathbf{x_i}) = 0$ for all i. The expected value of the residual μ_i given the set of regressors $\mathbf{x_i}$ is 0. This is the strict exogeneity assumption we discussed in Chapter 4 as it applies to cross-sectional data. It implies our model does not systematically overestimate the outcome $E[y_i - \hat{y}_i] > 0$ or underestimate the outcome $E[y_i - \hat{y}_i] < 0$ after we condition on the correct set of regressors $\mathbf{x_i}$.

CS1 and CS2 guarantee that OLS is unbiased.

The "conditional" part of the zero conditional mean assumption, CS2, causes some to express the assumption that the model is correctly specified. The model must include *all the relevant variables* and *no irrelevant variables*, but this description is somewhat ambiguous.[2] The inclusion of irrelevant variables only effects the efficiency of

[1] See Greene (2018) for applications of these tests and the relationships among the tests.

[2] This sort of description is typically used to simplify the discussion of the assumptions but this description does not suit our needs here (see Kellstedt and Whitten (2018)).

the estimator, it will not induce bias. The omission of an irrelevant variable will only cause one to systematically over-or-under estimate the effect of a variable of interest if two conditions are met.

To illustrate, imagine a PRF of the form depicted in Equation 1 that an analyst tries to model without one of the regressors, say x_{2i}. If the analyst estimates Equation 6, this systematic variation in the regressand associated with x_{2i} will now be reflected in the residual term $\hat{u}_i$.

$$y_i = \hat{\alpha}_0 + \hat{\beta}_1 x_{1i} + \hat{u}_i. \qquad \hat{u}_i = \hat{\beta}_2 x_{2i} + \hat{e}_i \tag{6}$$

Imagine further that there is a non-zero relationship between x_{1i} and x_{2i}, such that the relationship between the variables would be reflected by the following auxiliary regression model:

$$x_{1i} = \omega_0 + \gamma x_{2i} + \nu_i \tag{7}$$

The bias created by the omission of x_{2i} takes the form:

$$E[\hat{\beta}_1] = \beta_1 + \beta_2 \times \gamma \tag{8}$$

Equation 8 highlights the conditions where incorrect model specification creates bias. If x_{2i} belongs in the conditional model for y_i, *and* x_{2i} covaries with x_{1i}, then the omission of x_{2i} will bias the estimates of the parameter of interest, β_1. Changes in x_{1i} will not only cause changes in y_i, but x_{1i} will also vary systematically with the residual term $\hat{u}_i$ from Equation 6 as a consequences of the non-zero parameter γ from Equation 7, likely resulting in violations of the assumptions made about the model residuals, CS5 and CS6. This bias will not be eliminated as the sample size increases. If, on the other hand, either $\gamma = 0$ or $\beta_2 = 0$, the OLS estimator for β_1 will be unbiased.[3]

CS2 is also violated if y_i cause changes in x_i (reverse causality) or both y_i and x_i cause changes in each other (simultaneity). Measurement error in the regressors can also induce correlation between a regressor and the error.

CS3. No Serial Correlation: For $i \neq j$, $cov(\mu_i, \mu_j|\mathbf{x_i}) = 0$. The disturbance terms coinciding with μ_i and μ_j are uncorrelated conditional on the current value of the regressors. If this assumption is violated we say the residuals are autocorrelated or serially correlated.

CS4. Homoskedasticity: The variance of μ_i is constant conditioning only on the explanatory variables $var(\mu_i|\mathbf{x_i}) = \sigma^2$ for all i. If this assumption is violated, we say the residuals are heteroskedastic.

If either CS3 or CS4 are violated, our estimates of the model uncertainty, $\hat{\sigma}_e$, will be incorrect. This, in turn, will cause both our estimates of the variability of the individual coefficients (e.g. $\hat{\sigma}_{\hat{\beta}_1}$, $\hat{\sigma}_{\hat{\beta}_2}$, $\hat{\sigma}_{\hat{\alpha}_0}$) and the test statistics based on these estimates (e.g. $t_{\hat{\beta}_1}$, $t_{\hat{\beta}_2}$, $t_{\hat{\alpha}_1}$) to be incorrect. These violations create misleading inferences.[4]

In addition, there are two "minimal mathematical requirements" (Kellstedt and Whitten 2018, 211) for OLS. The first of these is no perfect collinearity: no regressor is constant or can be expressed as a linear function of other regressors. The second is that the number of observations must be greater than the number of regressors ($N > k$). While these are often stated as model assumptions (see: Gujarati (2003, 73)), they are not assumptions, as such, because their violation renders estimation literally impossible. One can estimate a model omitting a relevant variable x_{2i} but one cannot estimate a model without sufficient data. Software programs often drop "variables" from a model

[3]Borrowing from the literature on structural equation modeling, authors sometimes refer to x_{1i} as an "endogenous regressor" in this context. The intuition follows from the definition of weak exogeneity we outline in Chapter 4. If x_{1i} and x_{2i} are related, x_{2i} is part of the marginal model for x_{1i}. The covariation between x_2 and y_i means this marginal model cannot be ignored because x_{i1} is not determined outside of the system. We avoid that language here because, as we have mentioned previously, the zero-conditional mean assumption does not support hypothesis testing and statistical inference in the framework we outline below. To prevent confusion later, we don't want to refer to this as *the exogeneity assumption* here.

[4]Some violations are what we might call *pure* and occur independent of violations of the model assumptions we highlight above. In these cases, the violations represent a computational nuisance where variances and standard errors for the coefficients will be incorrect but the coefficient estimates remain unbiased. In such cases, standard practice dictates using OLS to estimate the coefficients and some robust estimator of the standard errors to correct for the violations. For example, Huber-White standard errors correct for heteroskedasticity or Newey-West standard errors correct for heteroskedasticity and autocorrelation. These solutions are inadequate when the violations occur as a consequence of the model misspecification errors highlighted above. As a practical matter, it is difficult to know, without further testing, whether violations of these assumptions are of the *pure* variety we have just described or are a consequence of some other model malady.

that do not have unique variation, e.g., if one regressor is either a perfect function of another or a perfect linear combination of other variables.

You will often see a normality assumption included in the Gauss-Markov assumption, but this assumption is not required. Many texts acknowledge that the normality assumption is not necessary for OLS to be BLUE but vaguely suggest that the assumption is necessary for hypothesis testing (see: Greene (2018, 23)). This isn't entirely accurate either. Lumley et al. (2002, 152) point out that the basis for hypothesis testing in linear regression, and in most contexts in applied statistics, is the central limit theorem. On average, the test statistics are normally distributed which allows us to make inferences using t-statistics and to construct confidence intervals for our estimates. The normality assumption is only required with extremely small samples, otherwise it is a sufficient but not necessary condition for statistical inference (Lumley et al. 2002, 157).[5] For our purposes, we can leave this assumption aside – if your sample size is so small that you wouldn't be able to conduct a hypothesis test without assuming normally distributed errors, you don't have enough data points to conduct a time series analysis.

Assumptions CS1-CS4 imply normality asymptotically and ensure that OLS is best, linear, unbiased, and efficient (BLUE). Thus, they are required for unbiased estimation and reliable inference. In applied settings, we typically do not view these assumptions as questions that can be left to faith. The challenges are the verification of these assumptions and the augmentation or correction of the OLS model or estimates to accomodate departures from these assumptions. A long literature develops, evaluates, and applies tests for violations of the assumptions, as well as methods to make OLS robust when they are violated. We will not review that literature here. Instead, we offer this basic exposition to highlight the standard OLS assumptions in the cross-sectional context so that we might, in the next section, better understand how these assumptions are often problematic in time series analysis.

2 Time Series Regression Models: The Assumptions

When we analyze time series data, we replace i with t in Equation 1, to denote that the variables are indexed by time rather than space:

$$y_t = \alpha_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + e_t. \tag{9}$$

This small change from the cross-sectional context introduces a host of new possibilities for inference and understanding. Like the linear regression model for cross-sectional data, time series regressions must satisfy a number of assumptions in order for the estimates to have desirable properties.

The central feature of time series data – namely that time series processes are not random – necessitates a statement of the Gauss-Markov assumptions that is separate from the assumptions outlined above. The assumptions in the previous section ensure that OLS is unbiased ane efficient (BLUE). These conditions are seldom met in time series regression and *cannot* be met in models that include lagged values of the dependent variable.

A weaker – and more realistic – set of assumptions guarantees the consistency of the OLS estimators. We will let $\mathbf{z_t}$ denote the variables and lags in the model. It can include: contemporaneous values of x_{jt}, where the subscript jt denotes one of $j = 1, ..., k$ explanatory variables at time t; lagged values of x_{jt}; lagged values of y_t; first-dfferences of y_t and x_{jt} (and their lags); as well as a constant, trend, and deterministic variables that are specified by the analyst to account for interventions.

TS1. Linearity: The joint data generating process for $\{(\mathbf{z_t}, y_t) : t = 1, 2, \ldots, T\}$ follows the linear model given by

$$y_t = \mathbf{z_t}\boldsymbol{\beta} + \mu_t \tag{10}$$

where $\{\mu_t : t = 1, 2, \ldots, T\}$ is the sequence of disturbances and T is the number of time periods.

TS2. Stationarity and Weak Dependence: The variables in the model are stationary and weakly dependent. While weak dependence is not a term we have introduced, it implies that the correlation between y_t and y_{t+h} (and x_{jt} and x_{t+h}) tends towards zero sufficiently quickly as $h \to \infty$. Stationary processes of the class we have considered are weakly dependent. We will relax this assumption in Chapter 6.

[5] Lumley et al. (2002, 152) show that repeated sampling from extremely skewed distributions still returns normally distributed statistics, even with sample sizes as small as 65. The threshold is likely much lower than the threshold in their example.

TS3. Zero Conditional Mean: $E(\mu_t|\mathbf{z_t}) = 0$ for $t = t$. This assumption implies that μ_t is contemporaneously uncorrelated with the regressors and thus that the μ_t are independent of the current values of the regressors but not necessarily the past and future values of the regressors. This assumption permits feedback from y_t to future values of the regressors, such that the explanatory variables may be correlated with past values of the disturbances. A variable that is weakly exogenous has a disturbance in period t that must be independent of current *and past* values of the explanatory variables. Thus, regressors that are weakly exogenous for the parameters in the model will satisfy the zero conditional mean assumption.

TS4. No serial correlation: For $t \neq s$, $cov(\mu_t, \mu_{t-s}|\mathbf{z_t}) = 0$. The disturbance terms coinciding with μ_t and μ_s are uncorrelated conditional on the current value of the regressors.

TS5. Homoscedasticity: The variance of μ_t is constant conditioning only on the current value of the regressors: $var(\mu_t|\mathbf{z_t}) = \sigma^2$ for all t.

Like with regression in the cross-sectional case, you must again meet the minimum mathematical requirements needed to estimate the model: no perfect collinearity and more observations than regressors ($T > k$).

Assumptions TS1-TS5 imply normality asymptotically. Thus, under the null hypothesis, each t-statistic has a t distribution, and each F-statistic has an F distribution. The usual construction of confidence intervals is also valid. However, OLS not BLUE. Instead it is consistent such that inferences from time series regression models are valid asymptotically. These assumptions require (only) weak exogeneity of x_{jt} for the parameters of interest in the model and thus permit the inclusion of lagged dependent variables. In practice, this is an important feature of the models that typically describe the dynamics of social science time series but the OLS estimators may be biased if x_{jt} is dynamic and TS4 is violated. Indeed, estimates of coefficients on lagged dependent variables may be significantly biased downward, and this increases with stronger autocorrelation in y_t.[6] The coefficients on the lagged values of the regressand will almost always be biased in small samples. This bias is eliminated as T increases, hence the consistency property of the OLS estimators for the parameters on $\mathbf{z_t}$.[7]

Under some conditions it may be possible to make stronger assumptions about the nature of exogeneity of the regressors, but only if lagged vaues of y_t do not belong in the model. In particular, OLS estimators will be BLUE if we can replace TS3, TS4, and TS5 with their stronger counterparts, TS3', TS4', and TS5'. TS3' denotes strict exogeneity, the orthogonality of μ_t and the regressors for *all* time periods. TS4' occurs when there is no serial correlation conditional on *past, present, and future* values of the regressors. In other words, knowing the error term at time t doesn't help predict the error in any other time period. And TS5' occurs when homoscedasticity exists for *all* t conditional on *past, present, and future* values of the regressors. These conditions rarely hold in time series regression models. So, instead, we focus on the weaker set of assumptions outlined above.

If assumptions TS1-TS5 are satisfied, OLS can be used for statistical inference given sufficient T. But the minor differences between the sets of assumptions outlined in the previous section and this section are not the only differences that you need to be aware of. When lagged values of the regressand or the regressors are included in the model for y_t, the basic correspondence between the estimated slope coefficients captures only the immediate, or short-run, effects of the explanatory variables. Some additional work is required to recover the total, or long-run, effects. Importantly, a host of new interpretational tools are available to you. In the next section we explain how these models can be specified and how the information recovered through estimation can be used to learn a great deal about the underlying relationships of interest.

3 A Strategy for Model Specification and Inference with Stationary Data

How do you approach model building and inference with time series data? In this chapter we assume that the outcome variable of interest is stationary (or trend stationary) and that the regressors are weakly exogenous for the

[6] See Keele and Kelly (2006) and Achen (2000) on this and other issues related to lagged dependent variables in regression models.

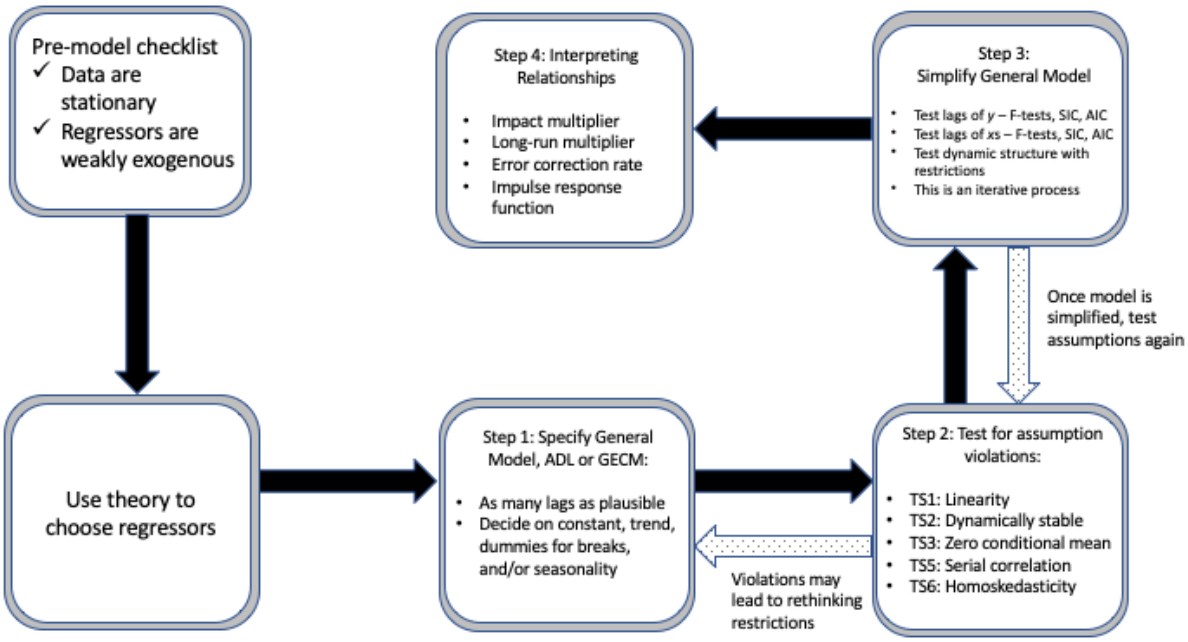
[7] Nickell (1981) discusses these types of biases in the context of panel data models with fixed effects. He refers to these small sample limitations of OLS estimators as Hurwicz-Type biases in the single-T case.

parameters of interest to you. It is of critical importance that the outcome variable does not contain a unit root. Unit root regressors must be handled cautiously, but can be included in the models we introduce in this chapter. We will have more to say about how to handle non-stationary regressors below.

We assume that before you begin to analyze the data, you have a theory describing the relationships between variables, but that it is unlikely that you have all the information you need to build a statistically adequate model based on theory alone. Our strategy will lead you to a final empirical specification based not just on theory but also on what you'll learn from the data. In part, this is because the error in dynamic regression models is "derived" – it depends on what is (and is not) included in the structural specification and on the sampling interval and the sample window.

We present a four step process designed to choose a suitable model for inference and map out the process in Figure 2. Step one is to select a general dynamic model specification that draws on theory to explain the present value of y_t. This involves determining which regressors to include, the highest reasonable lag length for the regressors, and the appropriate lagged values of y_t. The general model specification can be one of two flavors: the autoregressive distributed lag model (ADL) or the generalized error correction model (GECM). Each of these specifications encompass a wide range of dynamic structures. In step two, you check the statistical adequacy of the model with tests of whether the time series regression assumptions are met. In step three you test restrictions on the general model to produce a more parsimonious specification for inference. While the general model could be used for inference, imposing valid restrictions on it will result in more efficient estimates. Finally, in step four you draw inferences from the model by calculating quantities of interest. After describing each step, we illustrate the process with an example.

Figure 2: The Steps of Modeling with Stationary Data



4 Step 1: Specify A General Model

Step one in dynamic modeling is to choose a specification relating some dependent variable y_t to a set of explanatory variables as given by $\mathbf{z_t}$. $\mathbf{z_t}$ may include: contemporaneous and lagged values of x_{jt}; lagged values of y_t; first-dfferences of y_t and x_{jt} (and their lags); and a constant, trend, and deterministic variables that are specified by the

analyst to account for interventions. The specification we choose determines the structural relationship imposed on the data. In other words, it determines how y_t is allowed to respond to $\mathbf{Z}$ – at how many and which lags – and its own past behavior. Any dynamics not included in the structural model are left in the error. This must be avoided if you are to make reliable inferences from the estimated quantities. Our task is to select a specification sufficiently general as to encompass the underlying data generating process such that the model is dynamically complete, by which we mean there is no systematic dynamic structure in the resulting error.

Both theory and data are integral considerations in this task. Theory tells us which variables to include in the model, but theories are seldom specific enough to determine a particular dynamic structure, i.e. how many lags (of which variables) to include in the model. Often analysts admit uncertainty on this score. For example, in their investigation of judicial decisions, Mishler and Sheehan (1996, 176) note: "Although each hypothesis predicts a lag in the impact of public opinion on judicial decisions, none is sufficiently developed in theory to specify the precise length of the expected lag." Mondak and Smithey (1997) offer an equally telling note on the lack of specificity of theory as to the dynamic structure appropriate for models of Supreme Court behavior stating that they are "agnostic" about the specific dynamics of reasonable equilibrium behavior for the Supreme Court. Thus we recommend you start with a general model that allows for a variety of structural relationships and use the data to identify the appropriate dynamic structure (Hendry 1995; Beck 1991; De Boef and Keele 2008). Even when your theory suggests a particular dynamic structure, we recommend starting with a general model. Theory, after all, is always incomplete, always uncertain, and often wrong.

Two types of general model specifications are widely used in the social sciences: the autoregressive distributed lag model (ADL) and the generalized error correction model (GECM). Both specify a long-run equilibrium in y_t conditional on $\mathbf{z_t}$, where by equilibrium we mean that y_t tends to return to a long-run mean value over time. The two specifications entail the same dynamic structure, although they estimate different quantities directly. We consider each in turn.

4.1 The Autoregressive Distributed Lag Model (ADL)

The *autoregressive distributed lag model* (ADL) is the workhorse model for time series regression. The model specifies that y_t is a function of $\mathbf{z_t}$ where $\mathbf{z_t}$ contains some number of lags (p) of y_t and both contemporaneous and some number of lags (q) of the n explanatory variables, x_{jt}:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{n} \sum_{i=0}^{q} \beta_{ji} x_{jt-i} + \varepsilon_t$$

where ε_t is a white noise error satisfying the TS regression assumptions.. This specification requires that the regression assumptions given in Section 2 hold. The larger p and q, the more general the dynamic structure permitted in the ADL. If $p = 1$ and $q = 1$, the model is an ADL(1,1), with the first one indicating a single lagged y_t and the second one indicating one lag of x_t beyond the assumed estimation of a contemporaneous effect. The model specifies that one regressor has two direct effects on y_t through the contemporaneous value of the regressor, x_t, and the lag of that regressor, x_{t-1}, as well as indirect effects on y_t through y_{t-1}. There is one independent variable in the conditional relationship, but lags of x_t and y_t are necessary to capture the dynamics of the relationship.

The β_{ji}, are called *impact multipliers*. They tell us the immediate (or direct) effect of a one unit change in x_{jt} on y_t at lag zero and at each lag q. The inclusion of lagged values of y_t in the model causes a change in x_{jt} to exert an additional, indirect effect on y_t such that the total effect of a one unit change in x_j distributed over time is given by $\frac{\sum_{i=1}^{q} \beta_{ji}}{1-\sum_{i=1}^{p} \alpha_i}$. This quantity is called the *long-run multiplier* (LRM). We can express the *long-run equilibrium* as the steady state (static) solution to the model, which is given by the expected value of y_t given the expected value of an x_{jt}. Denoting the expected value by $*$ and since $E(\varepsilon_t) = 0$,

$$y^* = \alpha_0 + \sum_{i=1}^{p} \alpha_i y^* + \sum_{j=1}^{n} \sum_{i=0}^{q} \beta_{ji} x^*$$

$$y^* = \frac{\alpha_0 + \sum_{j=1}^{n} \sum_{i=0}^{q} \beta_{ji} x^*}{1 - \sum_{i=1}^{p} \alpha_i}.$$

Finally, $\sum_{i=1}^{p} \alpha_i - 1$ describe the pattern of adjustment in y_t when it is pushed from its equilibrium value either by a change in an x_{jt} or a shock in ε_t. When $p = 1$, $\alpha_1 - 1$ gives the *error correction rate*, the rate at which y_t changes to restore equilibrium. Finding a significant LRM with stationary data is evidence for the existence of a conditional stationary equilibrium. We will have much more to say about interpretation in Section 7.

4.2 The Generalized Error Correction Model (GECM)

The *generalized error correction model* (GECM) is a linear transformation of the ADL and thus while it has an alternative form, it implies exactly the same dynamic structure. The dependent variable in the GECM is the change in y_t (denoted Δy_t) and is a function of (short-run) changes in $\mathbf{x_t}$ ($\Delta \mathbf{x_t}$) and lagged values of y_t and $\mathbf{x_t}$. The main attraction of the GECM is that this particular re-parameterization of the ADL isolates the rate at which y_t changes to return to equilibrium after a change in $\mathbf{x_t}$ or any shocks (the error correction rate) and its standard error.[8]

To illustrate the equivalence of the ADL and GECM we begin with the ADL(1,1;1) where the third one indicates the inclusion of a single x_{jt} in the model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t. \tag{11}$$

Subtract y_{t-1} from both sides Equation 11:

$$\Delta y_t = \alpha_0 + (\alpha_1 - 1) y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t. \tag{12}$$

Add and subtract $\beta_0 x_{t-1}$ on the right side:

$$\Delta y_t = \alpha_0 + (\alpha_1 - 1) y_{t-1} + \beta_0 \Delta x_t + (\beta_0 + \beta_1) x_{t-1} + \varepsilon_t. \tag{13}$$

Collecting terms:

$$\Delta y_t = \alpha_0 + \alpha_1^* y_{t-1} + \beta_0^* \Delta x_t + \beta_1^* x_{t-1} + \varepsilon_t \tag{14}$$

where $\alpha_1^* = (\alpha_1 - 1)$, $\beta_0^* = \beta_0$, and $\beta_1^* = \beta_0 + \beta_1$ and the LRM is given by $-\frac{\beta_1^*}{\alpha_1^*}$ (De Boef and Keele 2008). This isomorphism highlights that we can retrieve the same information about short- and long-run effects from the GECM as from the ADL. Importantly, the fit of the model is identical to that of the ADL, although the R^2 will be different because the dependent variable is different. Similarly, the standard errors of comparable estimates will be the same.

More generally, the GECM corresponding to the ADL$(p, q; n)$ is given by:

$$\Delta y_t = \alpha_0 + \alpha_1^* y_{t-1} + \sum_{j=1}^{n} \beta_{j1}^* x_{jt-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} + \sum_{j=1}^{n} \beta_{j0}^* \Delta x_{jt} + \sum_{j=1}^{n} \sum_{i=1}^{q-1} \gamma_{ji}^* \Delta x_{jt-i} + \varepsilon_t \tag{15}$$

where $\alpha_1^* = \sum_{i=1}^{p} \alpha_i - 1$, $\beta_{j1}^* = \sum_{i=1}^{q} \beta_{ji}$, $\beta_{j0}^* = \beta_{j0}$, $\gamma_{j1}^* = \sum_{i=2}^{q} \beta_{ji}$, and for all $i > 1$, $\gamma_{ji}^* = \gamma_{ji-1}^* + \beta_{ji-1}$.

To see where the error correction model gets its name add and subtract $(\alpha_1 - 1) x_{t-1}$ on the right side of Equation 13 and regroup terms:

$$\Delta y_t = \alpha_0 + (\alpha_1 - 1)(y_{t-1} - x_{t-1}) + \beta_0 \Delta x_t + (\beta_0 + \beta_1 + \alpha_1 - 1) x_{t-1} + \varepsilon_t. \tag{16}$$

In this form it is easy to see that the difference between y_t and x_t in the previous period determines change in y_t in the current period (controlling for the other variables in the model). Note that while it appears as though the equilibrium relationship between x_t and y_t is one-to-one, the additional x_{t-1} on the right-hand side of the model allows the nature of the equilibrium relationship to be more general.

[8] An additional advantage of this parameterization is that Δx_{jt} and x_{jt-1} are not likely to be highly correlated and therefore hypothesis tests on the individual coefficients are more likely to be informative.

4.3 Step 1 in Practice

The choice of whether to work with the ADL or the GECM is largely a matter of taste but there are some considerations that might affect your decision. Most importantly, do you want to explain your dependent variable in levels or in differences? Other considerations are that the ADL is easier to pare down from the general model, even though it provides the same information as the GECM. Once you have chosen between the two, you have to make decisions about plausible lag length and deterministic terms.

4.3.1 How many lags should you start with in the general model?

The ADL and GECM are general models in the sense that they allow for a wide range of dynamic responses to a change in x_{jt}. But how many lags of y_t and x_{jt} (or Δy_t and Δx_{jt}) should we include in the model? Or, just how general should our specification of the structure be?

We should start with enough lags of y_t and x_t that, given our theoretical knowledge of the relationship and previous evidence, the existence of any additional significant lags would be a surprise. Generally this means a relatively short lag length for both y_t and x_t will suffice. The ACF and PACF for each time series also provide information as to the number of lags to include in the model. If, for example, the ACF and PACF for y_t suggest it is an AR(1) process, then it is likely that $p = 1$ will be sufficient to accommodate the dynamics in y_t. If the ACF and PACF suggests an AR(p) process, then $p = p$ should suffice. Similarly, if the ACF and PACF for each x_{jt} are consistent with an AR(1) process, then $q = 1$ is a reasonable maximum lag length, while for higher order AR(p) or an MA(q) process, additional lags may be necessary for the model to be general enough to capture the effects of the x_j. Best practices dictate that you begin with a more general lag structure that encompasses these expected dynamics. The specifications that appear to be best in preliminary analysis are only selected if the implied restrictions are supported by the data.

4.3.2 Dealing with Unit Root Regressors

These general models assume none of the variables in the model contain unit roots. It does not make sense for a unit root process to explain the behavior of a stationary variable. If the outcome is a function of a unit root process, then it, too, must contain a unit root. However, a stationary process may be a function of multiple unit root processes as long as the regressors are jointly stationary. This can be accomplished in a number of ways. We consider two possibilities here.

First, a unit root process can be made stationary by transforming the variable. You can do so by first-differencing the variable in question or you may impose some other transformation on the variable that renders it stationary, e.g., you might be able to recode levels of a time series in terms of growth rate to render it stationary. Transforming the variable changes the meaning of the regressor, but ensures hypothesis testing based on standard limiting distributions is valid. Second, including an even number of lags of a unit root regress makes the regressors jointly stationary. Think of it this way. To first-difference a variable is to subtract x_{t-1} from x_t. Thus, including both x_t and x_{t-1} as regressors accomplishes the same thing. In fact, when you include x_t and x_{t-1} as regressors and the variable contains a unit root, the coefficients on the two regressors will have oppositely signed and approximately equal coefficients.

You must proceed cautiously when including unit root regressors in levels. Only coefficients on stationary regressors have standard limiting distributions. Thus, if you include x_t and x_{t-1} in a model and that variable contains a unit root, t-tests on each x_t and x_{t-1} will not follow the t-distribution because each regressor individually contains a unit root. An F-test on these two regressors, however, will follow the F-distribution, because jointly the two regressors are stationary.

4.3.3 Dealing with Deterministic Trends, Structural Breaks, and Seasonality in the General Model

Both the ADL and GECM can also accommodate non-stationarity due to deterministic trends, and structural breaks in a straightforward manner. If y_t contains a linear deterministic trend, you include t as a regressor. The inclusion of

t means that you can interpret the quantities from the resulting model as if you had detrended the data first. If you suspect a structural break then intervention variables that account for the influence of the break should be included. A pulse intervention variable, where $Dp_t = 0$ prior to the break, $Dp_t = 1$ at the point of the break, and $Dp_t = 0$ in all subsequent periods, will accommodate the effect of a temporary shock. A step intervention variable, where $Ds_t = 0$ before the break and $Ds_t = 1$ at the point of the break and in all subsequent periods, will accommodate the effect of a shock that permanently altered the mean of y_t. A trend intervention variable, DT_t, can be used to model an event that caused a change in the slope of the trend. It is coded zero before the event and $DT_t = 1, 2, 3, \ldots T$ after the event. (See Chapter 3 for a more detailed discussion of intervention variables.) Of course, multiple interventions can be simultaneously modeled.

You can accommodate seasonality in the outcome variable in a similar fashion. You include variables in the regression to capture the seasonality. For example, if you have monthly data with $s = 12$, you include $DJan_t = 0$, coded zero unless the observation is in January, in which case it is coded as $DJan_t = 1$. You repeat this for each month of the year: $DFeb_t$, $DMar_t$, etc., for each of the 12 months of the year (and suppress the constant when estimating the model).

Wooldridge (2015, pp.337-338) suggests "it is a good idea to include a trend in the regression if any independent variable is trending, even if y_t is not." This allows for the possibility that movement around the trend in a given x_{jt} influences y_t. It is similarly good practice to include seasonal dummy variables if any x_{jt} is seasonal for the same reason. "Just as including a time trend in a regression has the interpretation of initially detrending the data, including seasonal dummies in a regression can be interpreted as deseasonalizing the data" (Wooldridge 2015, 341). The intuition relates to the formula for partial slope coefficients. The covariance between y_t and each x_{jt} is estimated less the covariances among the regressors. If any of the x_{jt} exhibit trend or seasonality, these dynamics will be partialed out of variability in that x_{jt} through the process of estimation. It should be noted, of course, that this covariation will manifest as multicollinearity among the x_{jt} and the trend or seasonal dummies, which could affect your inferences on x_{jt}.

5 Step 2: Test for Violations of the Time Series Regression Assumptions

After settling on a plausible general model specification, you next need to test the statistical adequacy of your model. In other words, you need to ask whether the regression assumptions are justified in the given specification. We focus our attention on testing for and handling violation of the time series regression assumptions of linearity (TS1), zero conditional mean (TS3), no serial correlation (TS4), and homoscedasticity (TS5). We have dealt extensively in Chapters 2 and 3 with the stationarity assumption (TS2). Here we briefly describe how to test fo dynamic stability in the regression context. We end with a few words on multi-collinearity and the normality of the regression disturbance.

5.1 TS1: Linearity

The linearity assumption presupposes that the model for the time series process y_t is constant and linear in its parameters, i.e., that the functional form of the relationship is correctly specified. When this assumption is violated, OLS estimators will be biased and inconsistent, and conventional inference procedures will be invalid. Violations of the assumption may occur because the underlying relationship is nonlinear or because the parameters are not constant.

Most models estimated in the social sciences assume linearity as a matter of course. But it is always possible that the true functional form for y_t is not linear.[9] For example, the relationship will be nonlinear if y_t grows or decays exponentially as a function of the explanatory variables or if the effects of the explanatory variables are believed to be multiplicative rather than additive – a small percentage change in one of the independent variables produces a

[9]Theory may also suggest that the dynamics of y_t differ based on its own previous values or the values of the independent variables. Threshold models can accommodate these dynamics. (see: Hansen (2000)). Theories that suggest the effects of $\mathbf{Z}$ on y_t depend on whether the process is in a given (unobserved)"regime." This can be modeled by allowing the parameters to take on different values in each of some fixed number of regimes (see: Hamilton (2010)).

proportional percentage change in the expected value of the dependent variable y_t. In the former case, y_t should be logged and in the latter case, both y_t and on regressor should be logged.

The estimates of the parameters describing the relationship between y_t and $\mathbf{z_t}$ will not be constant if we omit variables, including interventions or breaks, that condition the relationship between y_t and $\mathbf{z_t}$. For example, if the dynamic relationship is systematically different before and after an event, interventions should be added to the model.

Plots can be particularly helpful in diagnosing violations of TS1. Plots of the observed versus predicted values from the estimated model should be randomly distributed around a 45 degree line while plots of residuals versus predicted values (or values of individual x_{jt}) should be distributed randomly around zero with a constant variance. If the error terms are a systematic function of the predicted values of y_t, then there is something systematic in y_t that we need to account for before drawing inferences.

A number of formal tests for misspecification have been developed to detect nonlinearities. One of the most common is Ramsey's (1969) regression specification error test (RESET test), which provides a formal test for misspecification due to incorrect functional form and omitted variables, as well as correlation between $\mathbf{z_t}$ and μ_t. To conduct the test, we estimate an auxiliary regression of y_t on the squared (and possibly cubed and higher order functions) of the residuals from the original model specification. An F-test is used to evaluate the joint null hypothesis that the coefficients on the powers of fitted values are all zero. If the null hypothesis is rejected, it could be due to either non-linearity or an omitted variable. If the problem is non-linearity, logging the variables is likely a solution.

A large number of tests for parameter constancy have also been developed. In general, these tests follow a similar logic and structure in which the model is first fit on subsets of the data. Variations of the tests then generally assess whether recursive coefficient estimates are constant across samples or whether out-of-sample forecast errors are unbiased (have a mean of zero). We introduced empirical fluctuation tests for parameter constancy in Chapter 4. Tests for general parameter constancy may be informative about the constancy of all (Nyblom 1989) or particular parameters in the model (Brown, Durbin, and Evans 1975; Hansen 1992). Tests predicated on the assumption of a fixed number of breakpoints may assume either one or more known (Chow 1960) or unknown breakpoints (Quandt 1960; Kim and Siegmund 1989; Andrews 1993). If you identify a break point and can trace its origin to an exogenous event, then adding intervention variables will often solve the problem. However, it the parameter values vary widely, you will need to rethink the model specification more generally.

The results of these diagnostic tests may lead you to more interesting theoretical development and modeling exercises. Transforming your data so that a linear model fits is a statistical solution, but it may be worthwhile to think about what explains the non-linearity and then try to model that as well.

5.2 TS2: Stationarity and weak dependence

Tests for the stationarity of individual time series were discussed in Chapter 3. Generally, if the analyst has concluded that the outcome variable is stationary, the conditional model expressing y_t as a function of $\mathbf{Z}$ will be dynamically stable. It is, however, good practice to confirm that the estimated model is also dynamically stable, particularly if you are uncertain whether the outcome variable is stationary.

In models that contain a single lagged value of y_t, the condition for dynamic stability is simply that the absolute value of the coefficient on lagged y_t is less than one, $\hat{\alpha}_1 < 1$. The conditions for dynamic stability when additional lags of y_t are included in the model can be more difficult to determine. In general the roots of the characteristic equation must lie strictly outside the unit circle.[10] Many software packages report the roots of the characteristic equation such that violations of TS2 can be easily diagnosed. If you have estimated a model and the parameters are not dynamically stable – e.g. the coefficient of a single lagged dependent variable is either > 1 or < -1, or coefficients on multiple lagged dependent variables sum to > 1 or < -1 – you have probably incorrectly diagnosed y_t as stationary. Review that decision and consider the models covered in Chapter 6.

[10] Alternatively and equivalently, the inverse roots must lie strictly inside the unit circle.

5.3 TS3: Zero conditional mean

The zero conditional mean assumption states that $E(\mu_t|\mathbf{z_t}) = 0$. Omitted variables are often the source of violations of TS3. There are two types of omissions to consider. On the one hand, you may have omitted an x_{jt} that is correlated with y_t and at least one element of $\mathbf{z_t}$. As in the discussion in Section 1, these types of omission can cause you to over or underestimate parameters of interest. If you have the potential confounder on hand, or some variable that can proxy for the confounder, Ramsey's RESET test can be used to evaluate whether this is the source of the violation.

There is another sort of omission that can create bias that is particular to dynamic modeling. We often see violations of the zero-conditional mean assumption because we have misspecified $\mathbf{z_t}$, omitting necessary lags of y_t, lags of some x_{jt}, or both. Omitting necessary lags not only induces serial correlation in the residuals, leading to a violation of TS4, these omissions also produce omitted variable biases. If one follows the general-to-specific modeling procedure outlined in this chapter, these violations can typically be avoided.

Violations of weak exogeneity can also produce violations of the zero conditional mean assumption. If all the other regression assumptions are met, regressors that are weakly exogenous for the parameters of interest in the model are contemporaneously uncorrelated with the disturbance. Additionally, they are uncorrelated with *past* values of the disturbance. Tests for weak exogeneity are described in Chapter 4. If the weak exogeneity assumption is violated for a particular regressor of interest, you cannot test hypotheses or draw inferences about their parameters. If these parameters are of interest to you, no single-equation model specification is going to eliminate the problem. A multiple equation set-up is required. We discuss these models in Chapters 8 and 9. However, if you are not interested in drawing inferences about parameters on variables that are not weakly exogenous for those parameters then...

5.4 TS4: No Serial Correlation

Recall that assumption TS5 requires the disturbance to be serially uncorrelated conditional on the current value of the regressors: For $t \neq s$, $cov(\mu_t, \mu_{t-s}|\mathbf{z_t}) = 0$, such that the disturbance terms coinciding with μ_t and μ_s are uncorrelated conditional on the explanatory variables. In the presence of serially correlated errors, the estimated variances of the regression coefficients in OLS will be biased. In particular, in the presence of positively serially correlated errors, the OLS estimates of the standard errors will be smaller than the true standard errors. This increases the risk of type-I errors, rejecting a true null hypothesis.

Serially correlated errors may occur for a variety of reasons including (a) the omission of relevant explanatory variables, (b) unmodeled structural breaks, seasonality, or deterministic trends, (c) misspecification of the functional form, e.g., nonlinearities, or (d) dynamic misspecification, in particular the omission of relevant lags of the regressors. If tests reveal serial correlation in the errors, the analyst should consider each of these potential sources. The univariate analysis covered in Chapters 2 and 3 should help identify structural breaks, seasonality and deterministic trends.

In addition to these basic problems with hypotheis testing, serial correlation can induce bias in dynamic models. Keele and Kelly (2006, 191) demonstrate that OLS will underestimate the coefficients on the x_{jt} and overestimate the coefficients on y_{t-p} when the x_{jt} are dynamic and the residuals are autocorrelated. These biases are not eliminated at T increases. As such, one should never include lags of the regressand to "control for serial correlation," i.e., treat lags of y_t as nuisance parameters. Instead, one should consider why this serial correlation exists. If the autocorrelation in the residuals is a consequence of dynamic misspecification, as is often the case when theory alone is used to select a particular dynamic specification, one should include the correct number of lags to capture the omitted dynamics. If, on the other hand, the residual autocorrelation is not a consequence of erroneously omitting lags of x_{jt} or y_t, one should estimate an appropriate ARMA(p,q) model to capture these dynamics. Cook and Webb (2021) discuss a non-linear Wald test that can be used to arbitrate between these two cases.

In what follows, we will assume that you have modeled these features of the data, have not omitted relevant explanatory variables, and that the linear model is the correct functional form. As such, serial correlation is likely to be an indication of dynamic misspecification. In fact, the chief question for determining the statistical adequacy of a given dynamic regression is whether the model is dynamically complete. A model is free from dynamic misspecification if it includes enough lags of y_t and x_t (or Δy_t and Δx_t) that the inclusion of further lags provides no additional

information. A dynamically complete model must satisfy Assumption TS4. If it does not, then additional lags of y_t or x_{jt} (or Δy_t and Δx_{jt}) are needed to capture the dynamics of y_t (or Δy_t) – their omission will produce serially correlated disturbances and OLS estimates will be inconsistent.

There are multiple tests for serially correlated disturbances.[11]. The most widely used test is the Breusch-Godfrey Lagrange multiplier test (LM), which is asymptotically valid for specifications with lagged dependent variables (Breusch 1978; Breusch and Pagan 1980; Godfrey 1978).[12] We can test for serial correlation using the LM test as follows. First, estimate the general model of interest. Let's assume the analyst has estimated an ADL(1,1;2):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_{10} x_{1t} + \beta_{11} x_{1t-1} + \beta_{20} x_{2t} + \beta_{21} x_{2t-1} + \mu_t. \tag{17}$$

Next, estimate the auxiliary regression of $\hat{\mu}_t$ on the the regressors included in the original regression and h lagged residuals, where h denotes the number of lags for which we wish to test for serial correlation and $\hat{\mu}_t$ are the residuals from the original regression.

$$\hat{\mu}_t = \gamma_{00} + \pi_{11} y_{t-1} + \gamma_{10} x_{1t} + \gamma_{11} x_{1t-1} + \gamma_{20} x_{2t} + \gamma_{21} x_{2t-1} + \rho_1 \hat{\mu}_{t-1} + \rho_2 \hat{\mu}_{t-2} + \ldots + \rho_h \hat{\mu}_{t-h}. \tag{18}$$

If there is no serial correlation, then the coefficients on the lagged error terms must all be zero. The null hypothesis for the test is thus $\rho_1 = \rho_2 = \ldots = \rho_h = 0$. The alternative hypothesis is that $\rho_1 \neq 0$ and/or $\rho_2 \neq 0, \ldots,$ and/or $\rho_h \neq 0$, in which case the errors follow a general ARMA(p, q) process, where $h = max(p, q)$, so that the null hypothesis will be rejected if the residuals in the original equation follow an AR(p) or MA(q) or ARMA(p,q) process up to lag h. The LM test statistic is given by TR^2, where T is the number of observations in the auxiliary regression and R^2 is computed from the auxiliary regression. The LM test statistic will be larger as the explanatory power of the lagged residuals grows, in which case we are more likely to reject the null. The test statistic is distributed asymptotically as χ^2 with h degrees of freedom. Because the appropriate number of lags is not known, it is important to consider multiple choices for h.[13]

If we reject the null hypothesis of no serial correlation, then we have omitted important dynamics from the structural portion of the model and should add additional lags to the original specification. Looking at an ACF and PACF of the model residuals will provide diagnostic information of any omitted dynamic structure. For example, if the autocorrelations exhibit a pattern of decay, it is likely that additional lags of y_t and/or $\mathbf{x_t}$ need to be included in the model. The number of spikes in the PACF will suggest the number of lags that need to be added. Spikes in either the ACF or PACF at fixed intervals suggest the possibility of omitted seasonality. And a single spike may suggest an omitted intervention. After augmenting the original specification, you should retest for serial correlation in the new model, continuing until you arrive at a dynamically complete specification.[14]

5.5 TS5: Homoscedasticity

Assumption TS5 requires the variance of μ_t to be constant conditioning (only) on the explanatory variables at time t: $var(\mu_t|\mathbf{z_t}) = \sigma^2$ for all t. Heteroscedastic variance invalidates estimates of the standard errors on the estimated coefficients in OLS regressions.

[11] See Enders (2015); Pesaran (2015); Wooldridge (2015); Box-Steffensmeier et al. (2014) for additional tests

[12] If the regressors are strictly exogenous we can also test for first order serial correlation using the Durbin Watson d statistic or the LJung-Box Q test (Ljung and Box 1978). Neither test is appropriate in the presence of lagged y_t (Hayashi 2000). If the assumption is violated, the best course of action is generally to add one or more lags of x_t or y_t (or both) and proceed as above. One may also use generalized least squares to correct for serial correlation when the regressors are strictly exogenous, but because serial correlation is generally an indication of dynamic misspecification, GLS estimation is seldom used in time series analysis and will not be presented here. But see Wooldridge (2015).

[13] As a word of caution, note that the behavior of the LM test for serial correlation can be quite poor in small samples. The test also assumes that the residuals in the auxiliary regression are homoskedastic. If this assumption is violated, the size and power properties of the test suffer. Cumby and Huizinga (1992) discuss an alternative test that can be applied in these circumstances.

[14] One potential drawback of the LM test for serial correlation is that it is, in general, not valid in the presence of heteroscedastic disturbances. If heteroscedasticity is suspected, heteroscedastic-robust versions of the test should be used (Wooldridge 1991). The Cumby-Huizinga test of the null hypothesis that the regression error follows a moving average of order $q \geq 0$ against the alternative that autocorrelations are nonzero at lags greater than q. The test is robust to violations of the homoscedasticity assumption (Cumby and Huizinga 1990).

Violations of assumption TS5 indicate that the regression model does not predict y_t consistently across all values of y_t and may arise due to model misspecification – omitted variables, including trends and structural breaks, or incorrect functional form. Alternatively, heteroscedasticity may occur if measurement error in y_t lessens (or increases) over time. In particular, if measurement improves and measurement error decreases over time, we expect the error variance to decrease over time as well. Finally, it may be that the true data generating process itself is heteroscedastic, i.e., there are periods of low volatility and periods of high volatility in y_t. If these patterns are conditional – they are a function of the previous error variance – then a class of generalized autoregressive conditionally heteroscedastic (GARCH) models can be exploited for better prediction.[15]

There are a number of tests for heteroscedasticity. These tests make use of the fact that the variance of the residuals from the estimated model is a consistent estimator of the error variance (White et al. 1980). Under the null hypothesis of homoscedasticity, the error variance should not be a function of the explanatory variables in the model. As such, time series plots of regression residuals vs. predicted values (or individual explanatory variables) should not show any patterns. Patterns indicate not only the existence of heteroscedasticity, but also its potential form, which is useful information for identifying the appropriate form of tests for heteroscedasticty.

Tests for heteroscedasticity are typically based on the fit of auxiliary regressions in which the squared residuals are regressed on the explanatory variables (if the form of heteroscedasticity is assumed to be linear) and possibly functions of them (if the form of heteroscedasticity is assumed to be nonlinear). Two common forms of the test are given by Breusch and Pagan (1979) and White et al. (1980).[16]

The Breusch-Pagan test assumes normality of the data and that any heteroscedasticity is a linear function of the regressors. The test involves regressing the squared residuals on the original explanatory variables. Assume again we have estimated the ADL(1,1;2) model given in Equation 17. Estimate the auxiliary regression of $\hat{\mu}_t^2$ on the regressors included in the original regression.

$$\hat{\mu}_t^2 = \gamma_{00} + \pi_{11} y_{t-1} + \gamma_{10} x_{1t} + \gamma_{11} x_{1t-1} + \gamma_{20} x_{2t} + \gamma_{21} x_{2t-1} + v_t \tag{19}$$

$\hat{\mu}_t$ are the residuals from the original regression. The null hypothesis for the test is thus $\pi_{11} = \gamma_{10} = \ldots = \gamma_{21} = 0$. If the coefficient on any of the regressors is different from zero, then the variance is heteroscedastic. If they are jointly zero, R^2 of the model will be zero and the variance is (linearly) homoscedastic.

In contrast, White's test allows for a more general form of heteroscedasticity by regressing the squared residuals on all distinct regressors, their cross-products, and squares. Like the Bruesch-Pagan test, if the R^2 from this auxiliary regression is high, then we can explain the residual variance as a function of the independent variables and therefore the variance is heteroscedastic.

In both versions of the test, the regression can be used to compute a Lagrange multiplier test statistic given by TR^2, where T is the number of observations in the auxiliary regression and R^2 is computed from the auxiliary regression. The test statistic is distributed χ^2 with n degrees of freedom where n is the number of regressors in the auxiliary regression. As with the LM test for serial correlation, the LM test statistic will be larger as the explanatory power of the regressors grows, in which case we are more likely to reject the null hypothesis and conclude the disturbances are heteroscedastic. White's test is more likely to identify heteroscedasticity due to nonlinearities. However, the test uses a large number of degrees of freedom, reducing the power of the test in small samples. It has also been shown to have relatively lower power against conditional forms of heteroscedasticity.

It is important to note that it is possible to have both serial correlation and heteroscedasticity. Tests for heteroscedasticity are not robust to serial correlation as they assume independence of the errors. A reasonable course of action is thus to first test for serial correlation using a heteroscedastic robust test, correct for any serial correlation, and then test for heteroscedasticity.

If the null hypothesis is rejected using these or alternative tests, the typical solution is to apply heteroscedasticity robust standard errors (also known as White or Huber-White standard errors) to the original regression (White

[15]It may be plausible that heteroscedasticity is itself autoregressive. This does not present a violation of the time series regression assumptions, but by explicitly modeling volatility using GARCH models, estimates will be asymptotically more efficient than the OLS estimates (Engle 1982). Tests for ARCH effects and the large (and growing) variety of models for this type of data are covered in Enders (2015) and Wooldridge (2015). A well specified ADL or GECM model can be estimated as part of an ARCH-GARCH model of the heteroskedastic residuals.

[16]See Enders (2015); Pesaran (2015); Wooldridge (2015); Box-Steffensmeier et al. (2014) for additional tests.

et al. 1980; Huber 1967).[17] Heteroscedasticity robust standard errors use the squared residuals from the estimated regression model to estimate the error variances associated with each observation, which are then are used to produce consistent estimates of the standard errors. The use of heteroscedastic-consistent standard errors has become standard in applications where heteroscedasticity is suspected.

Newey and West (1987, 1994) extended the logic of heteroscedastic robust standard error estimation to also allow for violations of the no serial correlation assumption to produce heteroscedastic and autocorrelation consistent, or HAC, standard errors. The use of HAC standard errors is common in time series regressions with strictly exogenous regressors and they are applied in many unit root tests. They are not, however, frequently used in dynamic regressions with weakly exogenous regressors because serial correlation is generally treated as evidence of dynamic misspecification and because HAC standard errors are often so large as to render them useless. It is always possible, however, that the regression disturbances are serially correlated and heteroscedastic such that use of the combined heteroscedastic-autocorrelation-consistent, or HAC, standard errors may be a good strategy. In cases where TS4 and TS5 are not violated, the HAC standard errors will be equivalent to their OLS counterparts that assume serially uncorrelated and homoscedastic disturbances.

5.6 A Few Words on Normality

While the time series regression assumptions given in TS1-TS5 ensure that the OLS estimators are normally distributed asymptotically, in smaller samples ($T < 100$), it is good practice to test whether normality holds in sample. Non-normality may be caused by outliers or non-normality may be evidence that one or more of the other time series regression assumptions has been violated.

A normal probability plot or normal quantile plot (QQ plots) of the residuals provide visual evidence for and against normality. QQ plots present the quantiles of the data against those of a normal distribution with the same mean and variance as the sample data. If the sample data is normally distributed, the points will fall on the 45 degree line. Points that form a curve generally indicate excessively skewed data, while an s-shaped pattern indicates fatter tails than a normal distribution – there are too many or two few large errors in both directions – and thus excess kurtosis. If just a few points deviate from the 45 degree line, outliers may be the problem. Formal tests are also available, including the Jarque-Bera test. The test statistic measures the difference of the skewness and kurtosis of the series with those from the normal distribution.[18]

If there is significant evidence that the normality assumption does not hold for the model residuals, it is often a function of a significantly non-normal dependent variable or a violation of the linearity assumption. In this case transforming the data might fix both problems. Alternatively, and particularly in small samples, it may be that a few outliers drive the result, in which case it is important to determine if they are due to an omitted intervention that should be included in the model. Graphing the residuals should expose problematic outliers.

5.7 A Few Words on Multi-Collinearity

The mathematical requirement that there is no perfect collinearity states that no regressor is constant or can be expressed as a linear function of other regressors. OLS estimators are indeterminate in this case and standard errors are infinite. Such perfect collinearity is generally a data (or operator) error rather than an intrinsic feature of the data. However, because successive values of time series tend to be highly correlated, our explanatory variables are often highly and positively correlated.

The high collinearity among successive lags has important consequences. Their slopes will tend to be highly and negatively correlated. Further, because it is not possible to disentangle which variable is doing the work of explaining y_t, estimates can be imprecise, confidence intervals can be wide, and t-statistics will tend to be small in these circumstances. As such, individual hypothesis tests based on $t_{\hat{\beta}}$ may be unreliable. This is likely to be a problem with highly parameterized models in small to moderate samples, especially for the ADL.

[17] Weighted least squares can be applied if the form of the heteroscedasticity is known, but this solution is seldom applied since the development of heteroscedasticity robust standard errors.

[18] See Kennedy (2008) p. 78 on the Jarque-Bera test.

If you suspect that high levels of multicollinearity are affecting your inferences, you have a couple options. First, while the individual t-statistics can be rendered uninformative because of the high degree of correlation among the lags, F and χ^2 Wald statistics for restrictions on individual coefficients or groups of coefficients will not be effected. Your second option is to re-parameterize the model. Because an equivalent GECM models Δy_t as a function of single lag of x_{jt} and current and lagged values of Δx_{jt}, the regressors will tend to be nearly orthogonal. As a result, individual coefficients in the GECM will tend to be estimated with greater precision and hypothesis tests on individual estimates will tend to be more informative.

6 Step 3: Simplify the General Model

Having confirmed that your general model specification satisfies the time series regression assumptions, you have a statistically adequate general model that can be used for inference. However, it may be possible to impose restrictions on the dynamic structure and – when valid – estimate a more parsimonious model. Because time series processes tend to be autocorrelated, imposing restrictions on the general model usually reduces multicollinearity among the independent variables and thus increases the efficiency of estimates.

Given a statistically adequate general model, how should you approach simplifying the specification? For example, having started with an ADL(3,3;n) and having found no evidence the time series regression assumptions have been violated, how can you select a more parsimonious model? Two related issues are relevant. First, can higher order lags be dropped from the general specification? Second, what restrictions can be imposed on the form of the dynamic structure of the relationship between a given x_{jt} and y_t? The latter question amounts to asking whether the relationship between a given x_{jt} and y_t occurs at lag zero only, or at lag one only, or at lags zero and one, and so on for additional lags.

There is no set order for which of these decisions should come first. It is best to treat this as an iterative process, where the order can matter for which model you select as the final specification. It is a good idea to circle back and revisit decisions. If different strategies lead to different specifications, you can use model fit statistics to make comparisons and choose the model that best fits the data. As long as a model is dynamically complete and all the TS regression assumptions are met, reliable inferences can be drawn from the model. Other things equal, we would prefer a model that is parsimonious because estimates will be more precise. In what follows we guide you through the model specification process, focusing on the questions about model building that we posed above.

6.1 Can higher order lags be dropped from the general specification?

The starting point for model simplification begins by asking whether you can reduce p and q in the general model. Comparing nested specifications with F-tests and information criteria provide information, although t-tests may suffice in some cases. These tools help assess the trade-off between maximizing model fit – which by definition occurs in a more general model – versus a more parsimonious representation. You begin by estimating the general model and then iteratively test both the validity and dynamic completeness of successive restrictions. The process ends when further reductions in p and/or q result in serially correlated errors.

In this initial stage, the relevant comparison will be between a model with, for example, four lags of each variable to one with three lags, one with three lags to one with two lags, and so on. After making a determination as to the highest lag order, you can then consider restrictions on the form of the dynamic structure. Before we discuss these restricted forms, we want to offer some specific advice on some of tools that are most helpful in this initial stage, F-tests and information criteria.

6.1.1 F-Tests

The F-test allows you to test the validity of restrictions on a set of parameters in a given regression model. We estimate the general model specification and then a specification in which some parameters are restricted to zero. The F-statistic is used to test the null hypothesis that coefficients in the restricted model are all equal to zero.

For example, if in step one we selected an ADL(2,2,2), we might wish to test whether it is reasonable to reduce the highest lag of x_{1t} and x_{2t} from $q = 2$ to $q = 1$. We begin by estimating the unrestricted model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_{10} x_{1t} + \beta_{11} x_{1t-1} + \beta_{12} x_{1t-2} + \beta_{20} x_{2t} + \beta_{21} x_{2t-1} + \beta_{22} x_{2t-2} + \mu_t. \tag{20}$$

Then we estimated the restricted model given by

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_{10} x_{1t} + \beta_{11} x_{1t-1} + \beta_{20} x_{2t} + \beta_{21} x_{2t-1} + \mu_t. \tag{21}$$

We must take care to estimate both models over the same sample, T. Assuming both models are dynamically complete, we proceed by using the F-statistic to test the null hypothesis $\beta_{12} = \beta_{22} = 0$. (If the restricted model is not dynamically complete, the unrestricted model should be maintained.) The F-statistic is computed as:

$$F = \frac{(\frac{RSS_U - RSS_R}{k_R - k_U})}{(\frac{RSS_R}{T - k_R})} \tag{22}$$

where RSS_U is the residual sum of squares from estimation of the unrestricted model, RSS_R is the residual sum of squares from the estimation of the restricted model, T is the sample size, and k_U and k_R are the number of parameters estimated (including the constant) in the unrestricted and restricted models, respectively. In our example $k_U = 9$ and $k_R = 7$. Under the null hypothesis and assuming normally distributed errors, this statistic has an F-distribution with $(k_R - k_U, T - k_R)$ degrees of freedom. The restriction is binding if the F-statistic is sufficiently small, indicating a negligible difference in model fit by dropping $x_{1,t-2}$ and $x_{2,t-2}$ from the model.

6.1.2 Information Criteria

Information criteria are measures of model fit. The idea behind information criteria is to penalize improvement in model fit as a function of the loss in degrees of freedom that come from a more highly parameterized model. Information criteria differ to the extent that they penalize fit per estimated parameter.

The most commonly used information criterion is Akaike's information criterion (denoted AIC) and is given by $AIC = -2l/T + 2k/T$, where l is the log likelihood for the model, k is the number of estimated parameters, and T is the sample size.[19] A bias-corrected version of this criterion may be preferred in small samples. Schwartz's Bayesian information criterion (denoted as either SIC or SBC) is also widely used. The SIC imposes a stiffer penalty for lost degrees of freedom. As such, it will never select a less parsimonious model than the AIC. The Schwartz Bayesian information criterion is given by $SIC = -2l/T + (klogT)/T$. Whichever information criterion is used, the goal is to minimize its value. You estimate a set of potential specifications, calculate information criteria for each, and select the model with the information criterion that is closest to $-\infty$.

All models to be compared must be estimated on the same sample and all must be statistically adequate. In particular, dynamically incomplete models should not be selected, even if that model has the smallest information criterion. A minimum AIC or SIC is not a substitute for any of the formal tests for statistical adequacy. These criteria should only be used to make final comparisons among dynamically complete specifications. Both nested and non-nested models can be compared using information criteria as long as the models are fit to the same data. This feature of information criteria will be particularly useful when comparing model specifications that impose different types of restrictions. We turn to this task next.

6.2 What restrictions can be imposed on the form of the dynamic structure?

The restrictions we have covered so far determine the highest order lags to include in the general model. In this section we cover restrictions to the general model that impose alternative dynamic structures on the relationship between x_t and y_t. In other words, they constrain the *way* in which x_t affects y_t over time.

[19] An alternative formulation expresses the AIC in terms of the residual sum of squares from a model: $T \times log(\frac{\sum \hat{e}}{T}) + 2k$ or $e^{2k/T} \frac{\sum \hat{e}}{T}$ where T is the sample size and k is the number of estimated parameters (**?**, 63). **?**, 63 argue against the use of OLS residuals in the calculation, saying that the MLE estimate of the residual sum of squares should be used instead, but it is common to seethe AIC and SIC used to compare competing OLS specifications and it is not clear how or why this distinction is meaningful given that the measures are always used to compare the relative fit of competing models.

While a number of restrictions are possible, we describe five specifications likely to be relevant in applied work. We discuss the nature of the dynamic relationship implied by each, the restrictions imposed by each on both the ADL and GECM parameterizations of the general model, and the consequences of invalid restrictions for the estimation of the remaining parameters in the model. Then we identify potential strategies for how to approach selecting a final model for inference.

In what follows we discuss the model restrictions assuming the starting point is the ADL(1,1;1), but each specification is easily generalized to include additional independent variables and lags. Also note that each specification can be stated in terms of restrictions on the GECM. Following our exposition we provide a table summarizing the restrictions and the resulting models for both the ADL and GECM. We consider each restricted specification in turn.

6.2.1 The Partial Adjustment Model (PA)

The partial adjustment (PA) model specifies a contemporaneous relationship between x_t and y_t. As a restriction on the ADL(1,1;1), the PA model assumes $\beta_1 = 0$ and is given by:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \varepsilon_t.$$

The LRM in the PA model is $\frac{\beta_0}{1-\alpha_1}$. Thus the effects, like in the ADL, are distributed over time.

We can test this restriction using a t-test on β_1 in the ADL. If the restriction is imposed and is invalid, the coefficient estimates and the estimate of the LRM in the PA model will be biased. The degree and direction of bias is a function of the covariance of x_t and y_{t-1}, respectively, with x_{t-1}. If x_t is autocorrelated, the bias in β_0 in the ADL and β_0^* in the GECM will increase as a function of the autocorrelation in x_t. The LRM will typically be biased downward, so that the total effect of a change in x_t is underestimated.

6.2.2 The Dead Start Model

The dead start model, sometimes called a leading indicator model, specifies that x_t affects y_t only after a lag. Thus, the model imposes the restriction that $\beta_0 = 0$ on the ADL(1,1;1):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_{t-1} + \varepsilon_t.$$

The LRM for the dead start model is given by $\frac{\beta_1}{1-\alpha_1}$, implying, like the ADL and PA models, that the short-run effect is carried forward through time. We can test the restriction with a t-test on β_0 in the ADL. If the restriction is imposed when $\beta_0 \neq 0$, β_1 in the dead start model gives a biased estimate of the short-run effect and the LRM will be biased in similar fashion as for the PA model.

6.2.3 The Finite Distributed Lag Model (FDL)

The finite distributed lag (FDL) model specifies that x_t affects y_t with a lag (or lags), but x_t has only a direct effect on y_t; there is no indirect effect because we have excluded lagged values of y_t from the model. The FDL imposes the restriction that $\alpha_1 = 0$ in the ADL(1,1;1):

$$y_t = \alpha_0 + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t.$$

Thus, this specification implies that once we have controlled for x_t and x_{t-1}, no lags of y_t or additional lags of x_t affect current y_t. Because any effect of y_{t-1} is constrained to zero, the long-run effect is constrained to the sum of the impact multipliers: the LRM for x_t in the FDL is given by $\beta_0 + \beta_1$. If x_t is highly autocorrelated it will be difficult to get precise estimates of the separate effects of x_t at each lag in the FDL. However, the LRM still generally has desirable properties (Wooldridge 2015, 315). The restriction can be tested using a t-test of $\alpha_1 = 0$ in the ADL. If the restriction is invalid because $\alpha_1 \neq 0$, then β_1 will be biased and the LRM will be biased downward because $\beta_0 + \beta_1 < \frac{\beta_0+\beta_1}{1-\alpha_1}$.

6.2.4 The Static Model

The static model contains no lagged effects of y_t or x_t; the effect of x_t is instantaneous and limited to the current time point. The restriction imposed on the ADL(1,1;1) is $\alpha_1 = \beta_1 = 0$:

$$y_t = \alpha_0 + \beta_0 x_t + \varepsilon_t.$$

The restriction can be tested using an F-test. If the restriction imposes no significant loss of information and the static model is dynamically complete, the static model can be used for inference. The LRM for the static model is given by β_0 in the ADL, implying the long-run effect is equal to the short-run effect. Put differently, there is no dynamic effect of x_t on y_t.

The use of the term "static" comes from the fact that we are modeling a contemporaneous relationship between y_t and x_t, absent any dynamics. The specification says that, once the contemporaneous value of x_t has been controlled for, no lags of either y_t or x_t help to explain the current value of y_t. This is a strong requirement. If the restriction is invalid and $\alpha_1 \neq 0$ and/or $\beta_1 \neq 0$, then β_0 will capture neither short- nor long-run effects. This means the LRM will be biased downward unless x_t is a unit root process or $\beta_0\alpha_1 + \beta_1 = 0$. Further, we can expect the errors to be serially correlated as they will contain the omitted dynamics.

6.2.5 The First Differences Model

The first difference model restricts the ADL(1,1;1) such that $\alpha_1 = 1$ and $\beta_0 = -\beta_1$. With some reorganizing, this is:

$$\Delta y_t = \alpha_0 + \beta_0 \Delta x_t + \varepsilon_t.$$

The effect of x_t at time t in the first differences model is exactly counter-balanced by the effect at x_{t-1}. The LRM is undefined in the first difference specification; all effects are instantaneous and changes in y_t are unrelated to the levels of the variables. If the restriction is invalid we will have "thrown the baby out with the bathwater:" we can't even estimate a long-run relationship because it is assumed not to exist. The restriction can be tested by estimating the ADL and the first differences model and applying an F-test. This restriction will seldom be valid and should only be accepted if the residuals from the first differences model are not serially correlated.

While we lose a considerable amount of information about the long-run relationship by estimating the model in first differences, the specification is sufficient for basic hypothesis testing. If the LRM for x_t is non-zero in levels, the β_0 on Δx_t will be non-zero in the first-differences model. While this is sufficient to reject the null that y_t and x_t are unrelated, you should be careful not to over-interpret the results from this specification. That you have specified a model as though an LRM does not exist, does not mean that the LRM is actually undefined. That you are unable to calculate impulse response functions, does not mean that the process that produced the data is not dynamic.

6.3 Summary of Models

Table 1 presents each of these models and the implied restrictions associated with each, both in terms of the ADL (top half of the table) and the GECM (bottom half of the table). It is important to repeat that no restriction should be accepted as valid, regardless of the results of any t- or F-test, unless the restricted model is dynamically complete. Both the t- and F-tests tell us nothing about the statistical adequacy of the chosen model, only whether the fit of the restricted model is significantly different from that of the unrestricted model.

In applied work it is often useful to estimate multiple specifications, ensuring each is dynamically complete, and appeal to information criteria for model selection. This is particularly true when working from a higher order ADL or GECM where it is less obvious which restrictions should be tested in what order and where the order in which we conduct the tests can influence the final model selected. We provide guidance on how to approach this problem below.[20] As we will see, the substantive effects estimated from any dynamically complete model specification will

[20] Some statistical software packages automate the procedure by estimating *every* potential simplification of the general model, selecting a final model based on information criterion accompanied by a check for dynamic completeness in the selected model.

Table 1: The ADL, GECM, Restricted Versions of Each, and the Implied Restrictions

Model	Specification	Restriction on the ADL
ADL	$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$	None
Partial Adjustment	$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \varepsilon_t$	$\beta_1 = 0$
Dead Start	$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_{t-1} + \varepsilon_t$	$\beta_0 = 0$
Finite Distributed Lag	$y_t = \alpha_0 + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$	$\alpha_1 = 0$
Static	$y_t = \alpha_0 + \beta_0 x_t + \varepsilon_t$	$\alpha_1 = \beta_1 = 0$
First Differences	$\Delta y_t = \alpha_0 + \beta_0 \Delta x_t + \varepsilon_t$	$\alpha_1 = 1$, $\beta_0 = -\beta_1$

Model	Specification	Restriction on the GECM
GECM	$\Delta y_t = \alpha_0 + \alpha_1^* y_{t-1} + \beta_0^* \Delta x_t + \beta_1^* x_{t-1} + \varepsilon_t$	None
Partial Adjustment	$\Delta y_t = \alpha_0 + \alpha_1^* y_{t-1} + \beta_0^* x_t + \varepsilon_t$	$\beta_1^* = \beta_0^*$
Dead Start	$\Delta y_t = \alpha_0 + \alpha_1^* y_{t-1} + \beta_1^* x_{t-1} + \varepsilon_t$	$\beta_0^* = 0$
Finite Distributed Lag	$\Delta y_t = \alpha_0 - \alpha_1^* y_{t-1} + \beta_0^* \Delta x_t + \beta_1^* x_{t-1} + \varepsilon_t$	$\alpha_1^* = -1$
Static	$\Delta y_t = \alpha_0 - \alpha_1^* y_{t-1} + \beta_0^* \Delta x_t + \beta_1^* x_{t-1} + \varepsilon_t$	$\alpha_1^* = -1, \beta_1^* = \beta_0^*$
First Differences	$\Delta y_t = \alpha_0 + \beta_0^* \Delta x_t + \varepsilon_t$	$\alpha_1^* = 0, \beta_1^* = 0$

tend to be very similar. Differences, when they appear, are typically limited to the precision of the estimates and the dynamic paths the dependent variable takes in response to a shock to an x_{jt}.

7 Step 4: Interpreting Relationships

The purpose of estimating a dynamic regression is to characterize the relationships in the data, to test hypotheses about those relationships, and to use interesting quantities from the models to answer theoretically motivated questions. We might wish to know the effects of casualties on attitudes toward war. More specifically, we might want to know the immediate effect of war casualties on attitudes, whether any effects are ephemeral or long lasting, or the size of the total effect of some number of casualties. Similarly, we might be interested in characterizing attitudes toward immigration in the long-run: how do economic conditions, media frames, or events affect attitudes about immigration in the long-run; how quickly do attitudes deteriorate when economic conditions falter; or how long do we expect attitudes to be depressed following a downturn in the economy?

The answer to these and other questions about dynamic phenomena can be inferred from the estimates in any of the models we have considered. In what follows, we demonstrate how you can leverage information from the model you estimated to get a richer understanding of the underlying data generating process. For the sake of exposition, we will focus on a change in an individual x_{jt} (denoted simply x_t) on y_t. Though the various quantities of interest can be calculated for each element of $\mathbf{z_t}$, we typically reserve the computation of these quantities for the parameters of interest that describe the relationships between our variables of interest and our regressands.

7.0.1 What is the Direct Effect of a Unit Change in x_t on y_t?

Impact multipliers can be used to answer questions about the direct effects of changes in x_t. These may occur at any lag or lags and in the ADL, or restricted versions of the ADL. The impact multipliers are, simply, the estimated coefficients on contemporaneous x_t and each lag of x_t included in the model. Comparable information can be obtained from the GECM and restricted versions of the GECM by noting the equivalence of the estimated coefficients. For example, if we estimate an ADL, the direct effect of a one unit change in x_t at time t is given by the coefficient on x_t; the direct effect at $t-1$ is given by the coefficient on x_{t-1}. In the GECM, these same quantities are given by β_0^* and $\beta_1^* - \beta_0^*$ in the GECM. The combined direct effect of a unit change in x_t at time t and $t+1$, or $\beta_0 + \beta_1$ in the ADL, is given by β_1^* in the GECM.

7.0.2 What is the Total Effect of a Unit Change in x_t on y_t?

We have already provided the calculations for the total or long-run effect of a change in x_t on y_t in each of the specifications above. In the FDL and the static model, the total effect of a shock to x_t is equal to the (sum of the) direct effect(s). In these cases it is straightforward to calculate both the total effect and to determine whether if it is statistically significant using t- or F-tests. The first differences model does not allow for a long-run effect, it is presumed not to exist. However, in the dynamic specifications above, those including a lagged dependent variable, the LRM provides valuable information about total effect of a shock in x_t on y_t. Notably, the long-run effect can be substantially larger than the direct effect by virtue of a strongly autoregressive y_t, which carries the direct affect forward in time.

The LRM is not directly estimated in any of these model specifications and, thus, further work needs to be done to calculate its standard error. The problem is complicated by the fact that the LRM in these models is calculated from a ratio of coefficients and that there is no simple formula for the standard error of a ratio of coefficients.

There are two ways to retrieve the standard error. The first option is to use the delta method.[21] The second option is to estimate the following instrumental variables (IV) regression as suggested by **?**:

$$y_t = \phi_0 - \phi_1 \Delta y_t + \psi_0 x_t - \psi_1 \Delta x_t + \mu_t \tag{23}$$

where $\psi_1 = -\frac{\beta_1^*}{\alpha_1^*}$ and is the LRM. In fact, all the coefficients in Equation 23 are linear transformations of the GECM and ADL coefficients. The correspondence between this regression and the GECM in Equation 14 is given by: $\phi_0 = -\frac{\alpha_0}{\alpha_1^*}$, $\phi_1 = -\frac{\alpha_1^*+1}{\alpha_1^*}$, $\psi_0 = \beta_1^*$, and $\mu = -\frac{e}{\alpha_1^*}$. Translating from the ADL in Equation 11 we have: $\phi_0 = \eta\alpha_0$, $\phi_1 = \eta\alpha_1$, $\psi_0 = \eta(\beta_0 + \beta_1)$, $\psi_1 = \eta\beta_1$, $\mu = \eta\varepsilon_t$ and $\eta = \frac{1}{\alpha_1 - 1}$. A constant, current and lagged values of x_t and lagged y_t should be used as instruments for Δy_t to estimate the model (De Boef and Keele 2008). If a trend and/or interventions is included in the original model, a trend term and/or intervention variables should be included in the regression.

Restricted versions of the Bewley regression can be used to retrieve standard errors as well or these can be retrieved using the delta method (De Boef and Keele 2008). The partial adjustment version of the Bewley IV regression is:

$$y_t = \phi_0 - \phi_1 \Delta y_t + \psi_0 x_t + \mu_t \tag{24}$$

and the dead start version of the Bewley IV regression is:

$$y_t = \phi_0 - \phi_1 \Delta y_t + \psi_0 x_{t-1} + \mu_t \tag{25}$$

The static model and the finite distributed lag model do not have analogues to the Bewley regression because there are no dynamics in y_t.

It is important to recognize that just because the coefficient on lagged y_t (α_1 in the ADL parameterizations or α_1^* in the GECM parameterizations) is significant does not automatically imply a significant long-run relationship between y_t and any given x_{jt}, unless the (set of) lag(s) of x_t itself is also significant. However, even all lags of X_t are significant, the LRM will not be significant if the sum of the direct effects of x_t is zero.

7.0.3 How Quickly Does y_t Return to its Long-Run Equilibrium Value Following a Change in x_t?

While impact and long-run multipliers can tell us a great deal, examining the behavior of change over time – *how* y_t responds to changes in x_t – can tell us even more. The rate of error correction tells us the speed at which y_t adjusts to to any discrepancy in the equilibrium relationship between y_t and x_t in the previous period. As such, there is no error correction in the first differences model, which precludes equilibrium adjustment. Equilibrium adjustment occurs

[21] The delta method is used to approximate the standard error of a transformations of a random variable. The method uses a first-order Taylor series approximation. Because regression coefficients are themselves random variables, we can use this method to approximate the standard error of the LRM. In the first step the variance of the Taylor series approximation of the LRM about its mean is estimated. Then the variance of this approximation is used to estimate the variance of the LRM to arrive at the standard error.

upon impact in models with no lagged dependent variable. The error correction rate can, however, be calculated from any dynamic regression, i.e., any regression that includes a lagged dependent variable: the ADL, GECM, PA, and dead start models. In the ADL parameterizations it is given by $\alpha_1 - 1$ and in the GECM parameterizations by α_1^*.

The error correction rate will (typically) vary between 0 and -1 and must be negative.[22] Why must it be negative? The logic is most easily seen in the parameterization of the error correction model given in Equation 16 where the equilibrium term is given by $y_{t-1} - x_{t-1}$. Assume x_t increases. This creates a negative disequilibrium because the gap between y_t and x_t at time $t-1$ is too small. y_t must move in a positive direction to restore the equilibrium. But if x_t decreases, the disequilibrium is positive as $y_{t-1} - x_{t-1}$ is too big. In this case, restoring equilibrium requires a negative change in y_t. Thus error correction requires movement in the opposite direction from that of the disequilibrium and therefore must be negatively signed. If the error correction rate is -.50, then half of the disequilibrium is corrected at time t, half of the remaining disequilibrium is corrected at $t+1$, and so on. Larger error correction rates signify rapid adjustment in y_t, slower error correction rates signify more sluggish correction.

7.0.4 How Long Before Most of the Effect Plays Out? How Long Before it has Effectively Dissipated?

While the error correction rate tells us the rate of return to equilibrium following a change in x_t, it is often useful to illustrate the period-to-period changes in y_t that accompany a change in x_t using graphs of simple or cumulative lag distributions, also known as the simple and cumulative impulse response functions (IRFs). The former provides the estimated effect in a given period while the latter describes how the effect accumulates over time (until it reaches the LRM). The information in the lag distributions can be summarized by calculating the median lag length – how long it takes for half the long-run effect of a change in x_t to transpire – and mean lag length – how long it takes for the effect to effectively work its way out.

To illustrate, assume we have estimated an ADL(1,1;2) and that $\alpha_1 = .25$ and $\beta_{10} = 4$ and $\beta_{11} = 2$:

$$y_t = 0.25y_{t-1} + 4x_t + 2x_{t-1} + \varepsilon_t. \tag{26}$$

Now assume a temporary/one period shock of 1 to x_1 at $t = 0$; prior to and after the shock $x_1 = 0$. Assume $y_t - 1 = 0$ and that the value of x_2 did not change. Then ignoring x_2 we can calculate (and graph) the expected individual period effects as follows.

$$\begin{aligned}
\hat{y}_t &= .25(0) + 4(1) + 2(0) = 4 \\
\hat{y}_{t+1} &= .25(4) + 4(0) + 2(1) = .25(4) + 2 = 3 \\
\hat{y}_{t+2} &= .25\hat{y}_{t+1} + 4(0) + 2(0) = .25(3) = .75 \\
\hat{y}_{t+3} &= .25\hat{y}_{t+2} + 4(0) + 2(0) = .25(.75) = .1875 \\
\hat{y}_{t+4} &= .78\hat{y}_{t+3} + 4(0) + 2(0) = .25(.1875) = .046875...
\end{aligned}$$

Given the structure implied by the ADL, the direct effect of the one unit shock to x_t at time $t = 0$ is to increase the expected value of y_t at time t by β_0, here 4. The direct effect on y_t at time $t+1$ is given by β_1, or 2. There is an additional, indirect effect on y_t at time $t+1$ via y_{t-1} given by $\beta_0\alpha_1$. In fact, because y_{t-1} is in the model, the original change in x_t continues to affect y_t into the future. The effect of a shock in the ADL dissipates at a rate that is inversely proportional to $\sum_{i=1}^{p} \hat{\alpha}_i$.

The cumulative IRF is generated by summing the direct and indirect effects of a shock in all prior periods. The cumulative effect at time $t = 0$ is the value of the IRF at $t = 0$. At $t+1$ the effect is the sum of the time $t = 0$ effect and the time $t = t+1$ effect or, in our example, 7. The cumulative effect at time $t+2$ is the sum of the individual effects at $t = 0$, $t = t+1$ and $t = t+2$, which, in this example, is 7.75. As we will see in the examples below, it is particularly effective to present a graph of the lag distributions.

In order to determine the median and mean lag lengths, we standardize the individual period cumulative effects as a proportion of the LRM. The LRM for x_t in this example is given by $\frac{\beta_{01}-\beta_{11}}{1-\alpha_1} = \frac{2+4}{1-.25} = 8$. Clearly half

[22] Technically values between -1 and -2 are possible. However values in this range imply an oscillating return to equilibrium where y_t overcorrects at each step in the path to equilibrium. Typically, however, error correction rates in this range indicate serial correlation in the model errors.

of the effect is instantaneous, $4/8 = 0.5$, and by $t + 4$ most of the effect is gone ($7.98/8 = 0.998$) such that the mean lag length is 4.

The dynamic structures implied by the different models we have discussed are all different. Thus the lag distributions implied by a given model are unique. Figure 3 presents the simple (row one) and cumulative (row two) IRFs for the three models containing lagged values of the dependent variable: ADL, PA, and dead start models. The simple IRF for the ADL model in the in the left panel has direct effects at time $t = 0$ and $t = 1$ (given by $\beta_0 = 0.5$ and $\beta_1 = 0.3$). The presence of lagged y_t ($\alpha_1 = 0.8$) in the model means those effects continue to influence future values of y_t. The simple IRF for the PA model in the middle panel decays from the initial effect at time $t = 0$ (given by $\beta_0 = 0.5$) with no additional direct effect. Here again, the inclusion of y_{t-1} in the model propagates the effects of the shock into the future. For the dead start model in the right panel there is no effect at $t = 0$, the effect is lagged one period and thus the first nonzero value in the IRF occurs at time $t = t + 1$ and is equal to $\beta_1 = 0.3$. Then the IRF decays in a similar manner as that in the ADL and PA models.

The cumulative IRFs in the second row of the figure show the total effect of a one unit change in x_t in periods zero through 20. Thus, it sums the effects in the simple IRFs over time. Notably, the cumulative effects increase in diminishing fashion until the LRM is reached.

Figure 4 shows the simple and cumulative IRFs for the FDL, static, and first differences models. The simple IRF of the FDL (top left panel) shows the direct effects of each lag of x_t in the model (here two) and are given by the β_0 and β_1. But because the FDL does not contain lagged y_t, there are no additional effects on y_t. All changes are instantaneous in both the static (middle panel) and first differences (right panel) models and are given by the estimated coefficients on x_t or on Δx_t, respectively. The effects of a unit change in x_t in each of these models are not carried forward into the future. Therefore, the cumulative effects in each of these models are given by the sum of the direct effects.

Short median and mean lag lengths will occur in conjunction with fast error correction rates and describe a process with a dynamic structure the exhibit little inertia and respond quickly to change. A short median lag with a longer mean lag will likely occur with slower error correction and describe a process with a dynamic structure that quickly corrects a large part of disequilibrium but takes some time to complete its adjustment. Longer median and mean lag lengths will occur together with slower error correction when a process has considerable inertia.

The usefulness of any given inference will depend on the purpose underlying the analysis, but in general it is a good strategy to report all of the above information. It is also useful to discuss changes in the independent variables the size of which is meaningful. This may mean examining the effects of standard deviation or average changes in x_t rather than unit changes in x_t. In the next section we demonstrate how to calculate these quantities of interest from a model that was selected using the general-to-specific modeling procedure we outlined above.

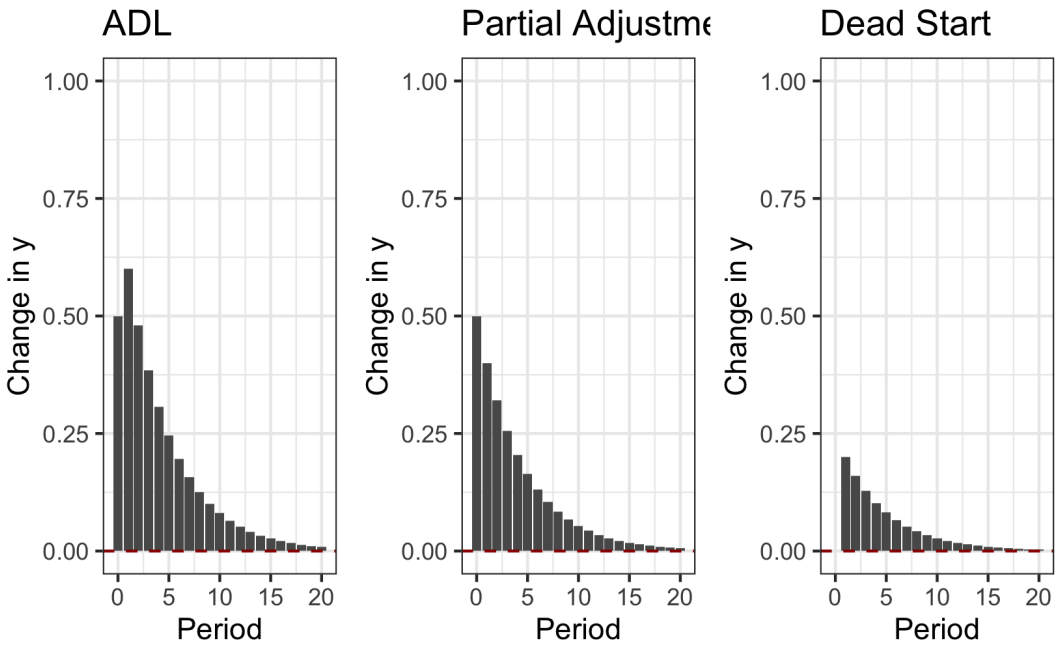
8 Examples

This section remains under construction.

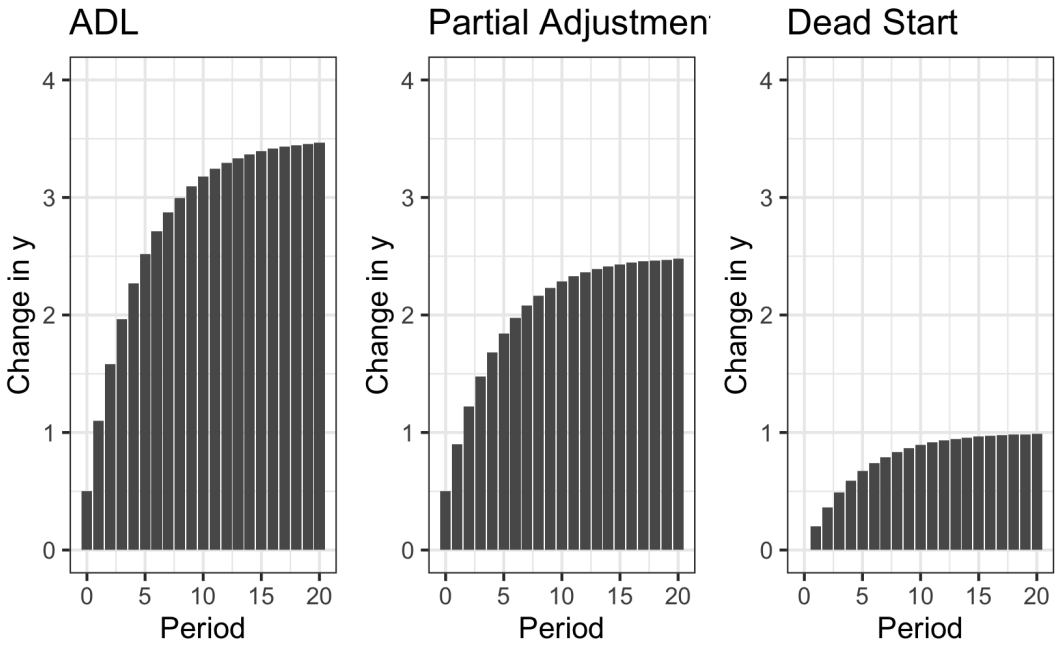
9 Conclusion

With stationary data and weakly exogenous regressors, the models of this chapter – the ADL and GECM and their variants – provide statistical tools to test hypotheses within models where a single regressand is posited as a function of one or more regressors. So long as everything is stationary, the worst pitfalls of time series analysis – equation imbalance, nonstandard distributions, and spurious regressions – are all avoidable. As we move to Chapter 6 and dealing with unit root data, these threats to inference loom large. The GECM we covered in this chapter is, again, suitable for such data but understanding its nuances remains the key to careful inferences.

Figure 3: Simple and Cumulative Impulse Response Functions for the ADL, Partial Adjustment, and Dead Start Models

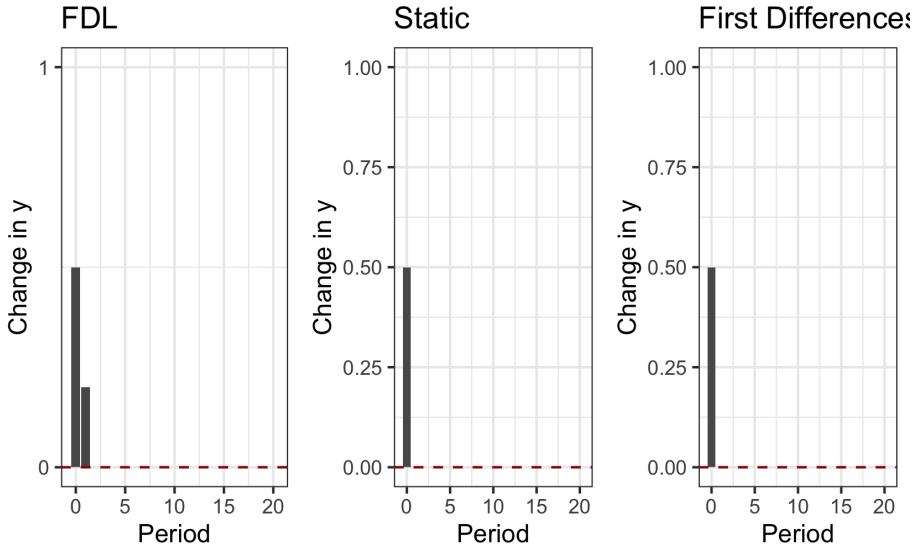


(a) Simple Impulse Response Function

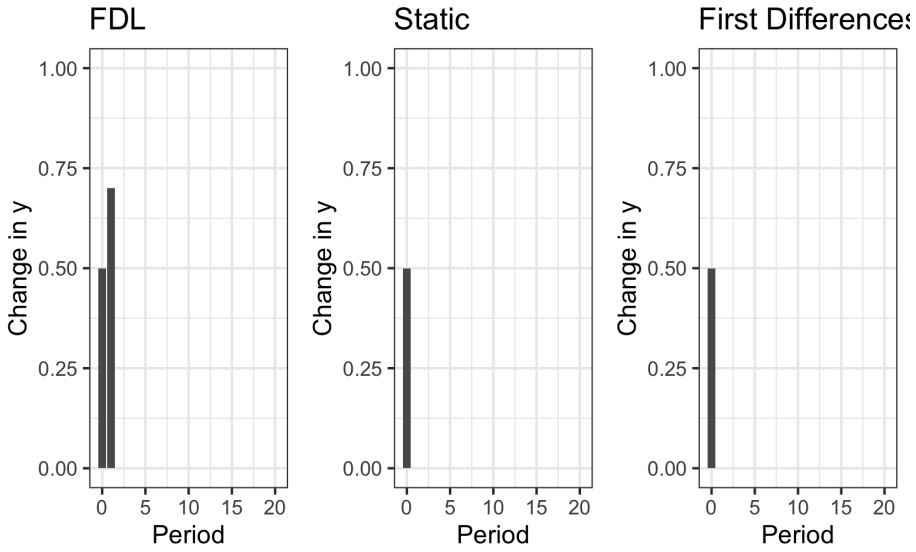


(b) Cumulative Impulse Response Function

Figure 4: Simple and Cumulative Impulse Response Functions for the FDL, Static, and First Differences Models



(a) Simple Impulse Response Function



(b) Cumulative Impulse Response Function

References

Achen, Christopher H. 2000. "Why Lagged Dependent Variables Can Supress the Explanatory Power of Other Independent Variables.".

Andrews, Donald WK. 1993. "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica: Journal of the Econometric Society* pp. 821–856.

Banerjee, Anindyaerjee, Anindya, Juan Dolado, John W. Galbraith, and David F. Hendry. 1993. *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data.* Oxford: Oxford University Press.

Beck, Nathaniel. 1991. "Model Selection: Are Time Series Techniques Useful in Cross-Sectional Problems?" 999: Midwest Meeting.

Box-Steffensmeier, Janet M, John R Freeman, Matthew P Hitt, and Jon CW Pevehouse. 2014. *Time series analysis for the social sciences.* Cambridge University Press.

Breusch, Trevor S. 1978. "Testing for autocorrelation in dynamic linear models." *Australian Economic Papers* 17(31): 334–355.

Breusch, Trevor S, and Adrian R Pagan. 1979. "A simple test for heteroscedasticity and random coefficient variation." *Econometrica: Journal of the Econometric Society* pp. 1287–1294.

Breusch, Trevor S, and Adrian R Pagan. 1980. "The Lagrange multiplier test and its applications to model specification in econometrics." *The review of economic studies* 47(1): 239–253.

Brown, Robert L, James Durbin, and James M Evans. 1975. "Techniques for Testing the Constancy of Regression Relationships over Time." *Journal of the Royal Statistical Society: Series B (Methodological)* 37(2): 149–163.

Chow, Gregory C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica: Journal of the Econometric Society* pp. 591–605.

Cook, Scott J, and Clayton Webb. 2021. "Lagged Outcomes, Lagged Predictors, and Lagged Errors: A Clarification on Common Factors." *Political Analysis* DOI: https://doi.org/10.1017/pan.2020.53: 1–9.

Cumby, Robert E, and John Huizinga. 1990. "Testing the Autocorrelation Structure of Disturbances in Ordinary Least Squares and Instrumental Variables Regressions.".

Cumby, Robert E., and John Huizinga. 1992. "Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions." *Econometrica* 60(1): 185–195.

De Boef, Suzanna, and Luke Keele. 2008. "Taking time seriously." *American Journal of Political Science* 52(1): 184–200.

Enders, Walter. 2015. *Applied Econometric Time Series.* 4th ed. New York: Wiley and Sons.

Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica: Journal of the Econometric Society* pp. 987–1007.

Godfrey, Leslie G. 1978. "Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables." *Econometrica: Journal of the Econometric Society* pp. 1303–1310.

Granger, C.W.J. 1991. *Modelling Economic Series: Readings in Econometric Methodology.* Advanced texts in econometrics Clarendon Press.

Greene, William H. 2018. *Econometric Analysis, 8th Edition.* New Jersey, USA: Pearson Education Press.

Gujarati, Damodar N. 2003. *Basic Econometrics.* International edition ed. New York: McGraw-Hill.

Hamilton, James D. 2010. "Macroeconometrics and Time Series Analysis." *In Durlauf S.N., Blume L.E. (eds) Macroeconometrics and Time Series Analysis.* pp. 202–209.

Hansen, Bruce E. 1992. "Testing for Parameter Instability in Linear Models." *Journal of policy Modeling* 14(4): 517–533.

Hansen, Bruce E. 2000. "Sample Splitting and Threshold Estimation." *Econometrica* 68(3): 575–603.

Hayashi, Fumio. 2000. "Econometrics." *Princeton University Press* pp. 142–144.

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

Huber, Peter J. 1967. The Behavior of Maximum Likelihood Estimates Under Non-standard Conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press pp. 221–233.

Keele, Luke J., and Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14(Spring): 186–205.

Kellstedt, Paul M, and Guy D Whitten. 2018. *The fundamentals of political science research*. Cambridge University Press.

Kennedy, Peter. 2008. *A Guide to Econometrics, 6th Edition*. Blackwell Publishing.

Kim, Hyune-Ju, and David Siegmund. 1989. "The Likelihood Ratio Test for a Change-point in Simple Linear Regression." *Biometrika* 76(3): 409–423.

Ljung, G. M., and G. E. P. Box. 1978. "On a measure of lack of fit in time series models." *Biometrika* 65(08): 297–303.

Lumley, Thomas, Paula Diehr, Scott Emerson, and Lu Chen. 2002. "The importance of the normality assumption in large public health data sets." *Annual review of public health* 23(1): 151–169.

Mishler, William, and Reginald S. Sheehan. 1996. "Public Opinion, the Attitudinal Model, and Supreme Court Decision Making: A Micro-Analytic Perspective." *The Journal of Politics* 58(February): 169–200.

Mondak, Jeffery J., and Shannon Ishiyama Smithey. 1997. "The Dynamics of Public Support for the Supreme Court." *The Journal of Politics* 59(November): 1114–1142.

Newey, Whitney K, and Kenneth D West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covaraince Matrix." *Econometrica* 55(3): 703–708.

Newey, Whitney K, and Kenneth D West. 1994. "Automatic Lag Selection in Covariance Matrix Estimation." *The Review of Economic Studies* 61(4): 631–653.

Nickell, Stephen. 1981. "Biases In Dynamic Models with Fixed Effects." *Econometrica* 49(November): 1417–26.

Nyblom, Jukka. 1989. "Testing for the Constancy of Parameters Over Time." *Journal of the American Statistical Association* 84(405): 223–230.

Pesaran, M Hashem. 2015. *Time Series and Panel Data Econometrics*. Oxford University Press.

Quandt, Richard. 1960. "ests of Hypotheses that a Linear System Obeys Two Separate Regime." *Journal of the American Statistical Association. American Statistical Association* 55.

Ramsey, James Bernard. 1969. "Tests for Specification Errors in lCassical Linear Least-Squares Regression Analysis." *Journal of the Royal Statistical Society: Series B (Methodological)* 31(2): 350–371.

White, Halbert et al. 1980. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *econometrica* 48(4): 817–838.

Wooldridge, Jeffrey M. 1991. "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances." *Journal of Econometrics* 47(1): 5–46.

Wooldridge, Jeffrey M. 2015. *Introductory econometrics: A modern approach*. Nelson Education.