Computational Economics **22:** 225–253, 2003.

Correcting for Omitted-Variable and Measurement-Error Bias in Autoregressive Model Estimation with Panel Data ⋆

P. A.V. B. SWAMY [1], I-LOK CHANG [2], JATINDER S. MEHTA [3] and GEORGE S. TAVLAS [4]

[1] *Statistical Methods Division, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics, Washington, DC 20212, U.S.A.;* [2] *Department of Mathematics and Statistics, The American University, Washington, DC 20016, U.S.A.;* [3] *Department of Mathematics, Temple University, Philadelphia, PA 19122, U.S.A.;* [4] *Director-Advisor, Economics Research Department, Bank of Greece, Athens, Greece*

(Accepted 29 January 2003)

Abstract. The parameter estimates based on an econometric equation are biased and can also be inconsistent when relevant regressors are omitted from the equation or when included regressors are measured with error. This problem gets complicated when the ‘true’ functional form of the equation is unknown. Here, we demonstrate how auxiliary variables, called concomitants, can be used to remove omitted-variable and measurement-error biases from the coefficients of an equation with the unknown ‘true’ functional form. The method is specifically designed for panel data. Numerical algorithms for enacting this procedure are presented and an illustration is given using a practical example of forecasting small-area employment from nonlinear autoregressive models.

Key words: autoregressive models, omitted-variable biases, measurement-error biases, concomitants, panel data

1. Introduction

The coefficient of a regressor in an econometric or a time-series model is free from bias when no relevant variable is omitted from the model, when included regressors are not measured with error, and when the ‘true’ functional form of the model is known. There are conceptual and practical obstacles to finding consistent estimators of bias-free coefficients, which include the biasing effects of omitted variables, measurement errors, and inaccuracies in the specified functional forms, and the practical difficulty of avoiding such biases. In this article, we demonstrate how, given suitable auxiliary variables called concomitants, omitted-variable and measurement-error biases can be removed even when the ‘true’ functional forms are unknown. Our method is consistent with that of Chang, Swamy, Hallahan and

⋆ The views expressed in this paper are the authors’ own and are not necessarily those of their respective institutions.

Tavlas (2000) and Swamy and Tavlas (2001) who do not deal with autoregressive models and panel data, which are the main focus in this paper.

We apply this approach to studies of the behavior over time of employment numbers for small domains, defined by, e.g., the intersections of Metropolitan Statistical Areas (MSAs) and Allocation Industry Codes (AICs) within each State of the U.S.

Section 2 considers a situation in which (i) the stochastic processes followed by economic variables are both neither stationary nor unit-root nonstationary, (ii) relevant regressors are omitted from the specified time series models, and (iii) variables included in these models are measured with error. It shows that, in this situation, the coefficients of autoregressive models are the sums of bias-free components and omitted-variable and measurement-error biases. The section also develops a consistent method for obtaining bias-corrected estimates of the coefficients of autoregressive models. Since a primary application of autoregressive models is forecasting, Section 3 proposes various measures for assessing the predictive accuracy of these models. Section 4 presents a consistent method for forecasting with autoregressive models when their coefficients contain omitted-variable and measurement-error biases. Section 5 uses this method to obtain the predictions of as yet unobserved values of small-area employment. Section 6 concludes.

2. Omitted Variables and Measurement Errors in Autoregressive Models

In this article, we are concerned exclusively with the problem of estimating employment totals for small domains, such as the intersections of States, MSAs, and AICs. There is need for such estimation because it can yield local-area unemployment estimates. State and local governments use local-area unemployment estimates to determine whether there is need for local employment and training services and programs. Local-area unemployment estimates are also used to determine the eligibility of an area for benefits in various Federal assistance programs. The U.S. Bureau of Labor Statistics (BLS) assumed technical responsibility for the program of computing such estimates. It uses a sample survey method to estimate total employment at the National and State levels. The criterion for this sample design is minimization of the variances of state-level estimators. Minimization of the variances of small-domain-level estimators is not done because it is prohibitively expensive. As a consequence, the sample sizes for small domains are very small if not zero.

Direct survey estimates of employment for a small area, based on data only from the sample units within the area, are not any more reliable than the estimates based on little or no sample information. To improve these estimates, we need to look for additional sources of information about employment for each and every small domain of interest. The Covered Employment and Wages program, generally known as the employment security 202 (ES202) program, is one such source. This program provides a virtual census of nonagricultural employees covered by State unemployment insurance laws and of civilian workers covered by the program of Unemployment Compensation for Federal Employees. In addition, it covers about 47% of all workers in agricultural industries. Even though the ES202 program is a comprehensive and accurate source of employment and wage data, by industry, at the national, state, and county levels, the problem is that ES202 data come in with a lag. These data for a given month are not available until after approximately 9 months. Because of this lag, estimates of small-area employment totals for the current quarter can be obtained only by averaging 10- to 12-month-ahead forecasts of those totals. For this reason, accurate 10- to 12-month-ahead forecasts need to be generated from the available ES202 data.

We cannot accurately model the behavior over time of small-area employment unless we understand this behavior very well. The findings of economists that inform us about such a behavior are: (i) the U.S. unemployment rate has a strong tendency to move counter-cyclically, upward in general business contractions and downward in expansions, (ii) this cyclical behavior is asymmetric in the sense that the unemployment rate increases at a faster rate than it decreases, and (iii) expansions tend to last longer than contractions. To these results, Montgomery, Zarnowitz, Tsay and Tiao (1998) add the results that (i) a linear univariate autoregressive integrated moving average model, denoted by ARIMA(p, d, q), or such a model with a multiplicative seasonal ARIMA factor, denoted by ARIMA$(p, d, q)(P, D, Q)$, failed to accurately represent the asymmetric cyclical behavior of the quarterly U.S. unemployment rate and (ii) a nonlinear model called the threshold autoregressive (TAR) model outperforms a seasonal univariate ARIMA model in terms of the mean squared error for multi-step-ahead forecasts during contractions or periods of rapidly increasing unemployment. They end their study with the remark that forecasting methods that fully exploit the non-linearity of employment or unemployment series remain to be developed. In this section, we explore the possibility that a generalized TAR model helps forecast small-area employment well in both economic contractions and expansions.

2.1. MONTHLY NONSTATIONARY AUTOREGRESSIVE MODELS

Let y^*_{gt} be the 'true' value of the employment for small area g and month t relative to its 'true' value in a base period. By 'true' value we mean the value that does not contain any measurement error. We assume that the process $\{y^*_{gt}\}$ is generated by a nonstationary moving average process of an order $Q_{gt} \le \infty(MA(Q_{gt}))$:

$$y^*_{gt} = \pi_{0gt} + \pi_{1gt}a_{gt} + \pi_{2gt}a_{g,t-1} + \cdots + \pi_{Q_{gt}+1,gt}a_{g,t-Q_{gt}}\,, \qquad (1)$$

where $\{a_{gt}\}$ is a white-noise process, and both the coefficients and the order are allowed to depend on g and t.

We also assume that the true functional form of Equation (1) exists but we do not know what it is. Allowing the coefficients of (1) to vary over time, we obtain a class of functional forms. Different patterns of variation in the πs generate different functional forms, and an unlimited number of functional forms of (1) can be generated by permitting the πs to vary freely. We do not know which one of these functional forms is 'true'. Moreover, the 'true' functional form of (1) may vary with g.

From Cramer's generalization of Wold's decomposition theorem for stationary processes (Greene, 2000, pp. 759–760), we know that Equation (1) represents a class of nonstationary processes. The intercept of (1) includes the deterministic component of this class. Given that we want (1) to capture the nonlinearities and nonstationarities inherent in the $\{y^*_{gt}\}$-process, these assumptions are reasonable.[1]

If any relevant variable is excluded from (1), then we can extend this equation to include it. Since we do not impose any specific functional form on (1) and do not exclude from it any relevant variable, it can cover the 'true' functional form as a special case, provided its coefficients are unrestricted. Therefore, we assume that for a particular pattern of variation in its coefficients, Equation (1) coincides with the 'true' model. We call such a pattern of variation the 'true' pattern.

We can rewrite Equation (1) in an autoregressive (AR) form as

$$\begin{aligned} y^*_{gt} &= \pi_{0gt} + \left(\frac{\pi_{1gt}a_{gt}}{y^*_{g,t-1}}\right) y^*_{g,t-1} + \left(\frac{\pi_{2gt}a_{g,t-1}}{y^*_{g,t-2}}\right) y^*_{g,t-2} + \cdots \\ &\quad + \left(\frac{\pi_{Q_{gt}+1,gt}a_{g,t-Q_{gt}}}{y^*_{g,t-Q_{gt}-1}}\right) y^*_{g,t-Q_{gt}-1} \\ &= \phi_{0gt} + \phi_{1gt}y^*_{g,t-1} + \phi_{2gt}y^*_{g,t-2} + \cdots + \phi_{Q_{gt}+1,gt}y^*_{g,t-Q_{gt}-1}. \end{aligned} \tag{2}$$

The biasing effects of omitted regressors, measurement errors, and of inaccuracies in the specified functional forms are known in econometrics. These effects are not present in the coefficients of Equation (2) because by construction, this equation does not have omitted or incorrectly measured variables and is not forced to follow a specific functional form. Hence, we call its coefficients with the 'true' pattern of variation 'the bias-free coefficients'. However, the problem with (2) is that its regressors are unlikely to be independent of its coefficients unless $\{y^*_{gt}\}$ is a stationary and invertible process, defined, e.g., in Greene (2000, pp. 752–754). Another problem is that the order, $Q_{gt} + 1$, is unknown. Consequently, the order of any specified AR process may not be equal to $Q_{gt} + 1$. Suppose that we specify the following AR model of order p:

$$y^*_{gt} = \phi^*_{0gt} + \phi^*_{1gt}y^*_{g,t-1} + \phi^*_{2gt}y^*_{g,t-2} + \cdots + \phi^*_{pgt}y^*_{g,t-p}. \tag{3}$$

If we are lucky, then p will be bigger than $Q_{gt} + 1$ for all t, in which case the first $Q_{gt} + 2$ coefficients of Equation (3) will be the same as those of Equation (2) and the remaining $p - Q_{gt} - 2$ coefficients will be zero. If we are unlucky, p will be smaller than $Q_{gt} + 1$ for all g and t. In this case, (3) differs from the 'true' model, a member of the class in (1), in that the lagged values, $y^*_{g,t-p-j}$, $j = 1$,

$2, \ldots, Q_{gt} - p + 1$, are omitted. Suppose that these omitted values are related to the lagged values in (3) according to the equation

$$\begin{aligned} y^*_{g,t-p-j} &= \beta_{0jgt} + \beta_{1jgt} y^*_{g,t-1} + \beta_{2jgt} y^*_{g,t-2} + \cdots + \beta_{pjgt} y^*_{g,t-p} \,. \\ &\quad (j = 1, 2, \ldots, Q_{gt} - p + 1). \end{aligned} \tag{4}$$

Substituting Equation (4) into Equation (2) gives

$$\begin{aligned} y^*_{gt} &= \left(\phi_{0gt} + \sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{0jgt}\right) + \left(\phi_{1gt} + \sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{1jgt}\right) y^*_{g,t-1} + \cdots \\ &\quad + \left(\phi_{pgt} + \sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{pjgt}\right) y^*_{g,t-p}\,, \end{aligned} \tag{5}$$

where each coefficient is equal to the corresponding coefficient in (3).

The coefficients of (5) have the invariance property that they are not altered when (2) is rewritten in terms of $(y^*_{g,t-1}, y^*_{g,t-2}, \ldots, y^*_{g,t-p})$ and a function of $(y^*_{g,t-1}, y^*_{g,t-2}, \ldots, y^*_{g,t-p})$ and $(y^*_{g,t-p-1}, y^*_{g,t-p-2}, \ldots, y^*_{g,t-Q_{gt}-1})$ (see Swamy et al., 1996). This result would not obtain if (4) were not used to derive Equation (5) from (2) (see Pratt and Schlaifer, 1984). The coefficients of (2) do not possess the invariance property. For this reason, the regressors of (2) that are omitted from (3) and the coefficients of Equation (2) are not unique, as shown by Pratt and Schlaifer (1984, p. 13). By contrast, the real-world relations are unique because they remain invariant against mere changes in the language we use to describe them (see Basmann, 1988, pp. 72–74). Thus, (5) shares this invariance property with the real-world relations.

We can get monthly time series data on employment for a number of small areas. Let the observations on the dependent variable of Equation (5) constructed from these data be denoted by y_{gt}, $g = 1, 2, \ldots, G$, $t = 1, 2, \ldots, T_g$. Suppose that these observations are measured with error and are equal to the sums of the 'true' values (y^*_{gt}) and measurement errors, denoted by v_{gt}, i.e., $y_{gt} = y^*_{gt} + v_{gt}$.[2] Substituting these observations into Equation (5) gives

$$\begin{aligned} y_{gt} &= \left(\phi_{0gt} + \sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{0jgt} + v_{gt}\right) + \\ &\quad + \left(\phi_{1gt} + \sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{1jgt}\right)\left(1 - \frac{v_{g,t-1}}{y_{g,t-1}}\right) y_{g,t-1} + \cdots \\ &\quad + \left(\phi_{pgt} + \sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{pjgt}\right)\left(1 - \frac{v_{g,t-p}}{y_{g,t-p}}\right) y_{g,t-p} \\ &= \gamma_{0gt} + \gamma_{1gt} y_{g,t-1} + \cdots + \gamma_{pgt} y_{g,t-p}\,, \end{aligned} \tag{6}$$

where

$$\gamma_{0gt} = \left(\phi_{0gt} + \sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{0jgt} + v_{gt}\right) \text{ and for } i = 1, 2, \ldots, p:$$

$$\gamma_{igt} = \phi_{igt} + \sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{ijgt} - \left(\phi_{igt} + \sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{ijgt}\right) \frac{v_{g,t-i}}{y_{g,t-i}}.$$

These definitions provide natural interpretations to the γs and any assumptions we might now make about them should not contradict these interpretations. We now provide these interpretations.

Interpretations of the coefficients of Equation (6): We call $(y^*_{g,t-1}, y^*_{g,t-2}, \ldots, y^*_{g,t-p})$ the included regressors because they are included in both models (2) and (3). We call $(y^*_{g,t-p-1}, y^*_{g,t-p-2}, \ldots, y^*_{g,t-Q_{gt}-1})$ omitted regressors because they are included in model (2) but not in model (3). It can be seen that the intercepts of equations in (4) are the portions of the omitted regressors remaining after the effects of the included regressors have been removed. Therefore, an interpretation of the intercept of (6) is that it is the sum of the bias-free intercept of model (2), a combination of the portions, β_{0jgt}, of omitted regressors having the coefficients, $\phi_{p+j,gt}$, on omitted regressors as its coefficients, and the measurement error in the dependent variable of (6). The coefficient on the i-th non-constant regressor of (6) is also the sum of three terms, ϕ_{igt}, $\sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{ijgt}$, and $(\phi_{igt} + \sum_{j=1}^{Q_{gt}-p+1} \phi_{p+j,gt}\beta_{ijgt})(-\frac{v_{g,t-i}}{y_{g,t-i}})$, which have the following interpretations: The first term with the 'true' pattern of variation is the bias-free coefficient on the i-th non-constant regressor of Equation (2). The second term represents the indirect effect due to the fact that the 'true' value of the i-th included regressor affects the 'true' values of omitted regressors (Equation (4)) that, in turn, affect the 'true' value of the dependent variable (Equation (2)). This term is called 'omitted-variables bias'. The coefficients of (6) do not contain the same magnitude of omitted-variables bias because the coefficients of (4) are not equal to each other. Finally, the third term captures the effect of incorrectly measuring the i-th included regressor. It is called 'measurement-error bias'. This bias is also not the same for different coefficients in (6). We call the coefficients of (6) the 'impure' coefficients because they are contaminated by omitted-variable and measurement-error biases. If these biases are present, as they usually are, then the coefficients of (6) may not be constant and may have the real-world sources of variation.

The real-world sources of variation in the coefficients of Equation (6): The first term of each of the coefficients of (6) is not constant if the 'true' Equation in (2) is not linear. We have already pointed out that a model that accurately represents the asymmetric cyclical behavior of employment series cannot be linear. The second term of each of the coefficients of (6) is also not constant if Equations (2) and (4) are not linear and the set of omitted variables changes over time. Under all these conditions and the variations in the ratios of measurement errors to the corresponding included regressors, the third term of each of the coefficients of (6) is a variable.

We have already pointed out that in Equation (2), the regressors can be correlated with their own coefficients. These types of correlations are usually present in linear-in-variables representations of nonlinear equations (see, e.g., Narasimham et al., 1988). The correlations between the regressors of (2) and their coefficients get strengthened further as we go from (2) to (6). Note that the third term of the coefficient on the i-th non-constant regressor in (6) involves the regressor itself. For this reason, the third terms of the coefficients of (6) are the additional factors strengthening the correlations between the coefficients and the regressors of (6) further. These coefficients are also not independent of each other because they have a common source of variation from the coefficients, $\phi_{p+j,gt}$, on omitted regressors. The implication of the lack of independence between the regressors and the coefficients of (6) is that the conditional expectation of y_{gt} given its lagged values included in (6) is not just a linear function of the conditioning variables. Any tests of hypotheses about the 'true' order and the coefficients of Equation (2) can give false conclusions if they are based on such a linear function.

2.2. SEPARATION OF BIAS-FREE COEFFICIENTS FROM OMITTED-VARIABLE AND MEASUREMENT-ERROR BIASES

One question that needs to be answered before estimating (6) is that of parametrization: which features of Equation (6) ought to be treated as constant parameters? While the correct answer is unknown, it is clear that Equation (6) suffers from internal inconsistencies if we adopt a parametrization that is not consistent with the interpretations of the coefficients of (6) given above. Under certain conditions to be specified below, the following assumption avoids such inconsistencies.

ASSUMPTION 1. *(i) The coefficients of Equation (6) satisfy the stochastic equations*

$$\gamma_{igt} = \bar{\gamma}_i + \sum_{h=1}^{K-1} \alpha_{ih} z_{hgt} + \mu_{ig} + \sum_{\ell=1}^{m} l_{ie} \varepsilon_{\ell gt}, \quad (i = 0, 1, \dots, p, g = 1, 2, \dots, G, t = 1, 2, \dots, T_g), \tag{7}$$

where the z_{hgt} are called the 'concomitants', no z_{hgt} is equal to 1 for all t, and the $l_{i\ell}$ are known constants;

(ii) The m-vector $\varepsilon_{gt} = (\varepsilon_{1gt}, \varepsilon_{2gt}, \dots, \varepsilon_{mgt})'$ follows the stochastic equation

$$\varepsilon_{gt} = \varphi_g \varepsilon_{g,t-1} + a_{gt}, \tag{8}$$

where φ_g is an $m \times m$ matrix, $a_{gt} = (a_{1gt}, a_{2gt}, \ldots, a_{mgt})'$ is distributed with $Ea_{gt} = 0$ and

$$Ea_{gt}a'_{g't'} = \begin{cases} \sigma_g^2 \Delta_g \text{ if } g = g' \text{ and } t = t' \\ 0 \text{ if } g \neq g' \text{ and } t \neq t' \end{cases} ;$$

(iii) The $(p+1)$-vector $\mu_g = (\mu_{0g}, \mu_{1g}, \ldots, \mu_{pg})'$ is distributed with $E\mu_g = 0$, and

$$E\mu_g\mu'_{g'} = \begin{cases} \Delta \text{ if } g = g' \\ 0 \text{ if } g \neq g' \end{cases} ;$$

(iv) The vectors μ_g and ε_{gt} are independent of each other for all g and t and each of them varies independently across g.

The concomitants are introduced to explain the variation in the coefficients of Equation (6). The greater the proportion of the variation explained by the concomitants the better. In (7), the term μ_{ig} is constant through time; it is an attribute of small area g that is unaccounted for by the included concomitants. The term $\varepsilon_{\ell gt}$ differs among small areas both at a point in time and through time; it is a portion of γ_{igt} that is not explained by the other terms on the right-hand side of (7). Recall that the omitted-variables and measurement-error bias contained in γ_{igt} is equal to the difference between γ_{igt} and ϕ_{igt} in (2). We can assume that the first term on the right-hand side of Equation (7) is equal to ϕ_{igt}, unless the process $\{y^*_{gt}\}$ is non-stationary or nonlinear, in which case ϕ_{igt} can be equal to $\bar{\gamma}_i + \sum_{h \in P_i} \alpha_{ih} z_{hgt}$, where $P_i \subset P = \{1, 2, \ldots, K-1\}$, which is the 'index set' formed by the values of h. Recall that ϕ_{igt} with the 'true' pattern of variation is a bias-free coefficient and an estimate of $\bar{\gamma}_i + \sum_{h \in P_i} \alpha_{ih} z_{hgt}$ is treated as a bias-corrected estimate of ϕ_{igt}. Judgments about the goodness of the fit of Equation (2) to given data should be based on the estimates of the bias-free terms of the coefficients of Equation (6), but not on the estimates of its impure coefficients. Estimation of the fixed coefficients and the error terms of Equation (7) will be discussed in Section 4 below. The coefficients $l_{i\ell}$ are included in Equation (7) to give the user the option of not including an error term in some coefficients of Equation (6). If $l_{i\ell}$ is zero for all ℓ, then γ_{igt} does not have an error term other than μ_{ig}. The number of error terms in Equation (7) is $(p+1)(m+1)$. This number can be reduced to $2(p+1)$ by setting (i) $p+1 = m$ and (ii) $l_{i\ell} = 1$ if $i = \ell$ and $= 0$ if $i \neq \ell$. We usually impose these restrictions. We also usually restrict φ_g in Equation (8) to be diagonal. This does not mean that a non-diagonal φ_g is never appropriate. We need to use a Monte Carlo algorithm developed in Chang et al. (2000) to achieve convergence of an iterative estimation method we apply to Equations (6)–(8) when φ_g is non-diagonal. This algorithm is very time consuming. We use a quadratic programming algorithm to estimate diagonal φ_g matrices under the restrictions that their diagonal elements lie between –1 and 1 (Chang et al., 2000). This algorithm generally converges.

Assumption 1 gives a parametrization of Equation (6) and the conditional moments of y_{gt}, given the values of the concomitants, if Equation (7) includes only those concomitants that satisfy the following assumption.

ASSUMPTION 2. *Given the values of the concomitants in (7), the regressors of (6) are independent of the μ_{ig} and $\varepsilon_{\ell gt}$ in (7) for all g and t.*

To understand this assumption, note that the decompositions of the γ_{igt}s in (7), unlike those in Equation (6), depend on our choice of concomitants. Assumption 2 incorporates the idea that the regressors of (6) can be conditionally independent of the μ_{ig} and $\varepsilon_{\ell gt}$, given the values of the concomitants, even though they are not unconditionally independent of their coefficients. What Assumption 2 really says is that the regressors of (6) are correlated with their coefficients because of the first K terms on the right-hand side of Equation (7), but once these terms are subtracted from the coefficient on the left-hand side of Equation (7), the error terms on the right-hand side of Equation (7) are independent of the regressors of (6). Thus, by using a decomposition of the coefficients of (6) in (7) that is different from that given right below (6), we find a solution to the problem of correlation between the regressors of (6) and their coefficients. If the decompositions of the coefficients of (6) given in (7) were the same as those given right below (6), then Assumption 2 would be false.

2.3. CONDITIONAL EXPECTATIONS IN THE PRESENCE OF OMITTED-VARIABLE AND MEASUREMENT-ERROR BIASES

Substituting (7) into (6) gives an equation whose fixed coefficients can be consistently estimated:

$$
\begin{aligned}
y_{gt} \;=\; & \bar{\gamma}_0 + \bar{\gamma}_1 y_{g,t-1} + \cdots + \bar{\gamma}_p y_{g,t-p} + \sum_{h=1}^{K-1} \alpha_{0h} z_{hgt} + \\
& + \sum_{h=1}^{K-1} \alpha_{1h} z_{hgt} y_{g,t-1} + \cdots + \sum_{h=1}^{K-1} \alpha_{ph} z_{hgt} y_{g,t-p} \\
& \mu_{0g} + \sum_{\ell=1}^{m} l_{0\ell}\varepsilon_{\ell gt} + \left(\mu_{1g} + \sum_{\ell=1}^{m} l_{1\ell}\varepsilon_{\ell gt}\right) y_{g,t-1} + \cdots \\
& + \left(\mu_{pg} + \sum_{\ell=1}^{m} l_{p\ell}\varepsilon_{\ell gt}\right) y_{g,t-p} \\
& (g = 1, 2, \ldots, G, t = 1, 2, \ldots, T_g).
\end{aligned} \tag{9}
$$

This is a nonlinear AR model whose desirability is that it is derived from (6) without contradicting the interpretations of the coefficients of (6). The error terms in (9) are both heteroscedastic and serially correlated. Our derivation of Equation (9) is justified by its producing such errors with no appeal to any arbitrary heteroscedasticity assumption.

Under Assumptions 1 and 2, the right-hand side of (9) with the error terms suppressed gives the conditional expectation of y_{gt} as a nonlinear function of the conditioning variables. Equation (9) is a generalized TAR model if one of its z_{hgt} is equal to the logarithm of $y_{g,t-2}/y_{g,t-3}$. This can be compared with the Montgomery et al. (1998, p. 482) type TAR model for the first difference of y_{gt}, which is an AR model of order 2 with only two distinct sets of values for its coefficients. One of these two sets is obtained if $y_{g,t-2} - y_{g,t-3} \leq 0.1$ and the other is obtained otherwise. Note that exact multicollinearity results if $y_{g,t-2} - y_{g,t-3}$ is used as one of the z_{hgt} in (9).

3. Assessing the Predictive Accuracy of Equation (9)

By changing the set of concomitants in (7) and/or changing the value of p we can derive different versions of Equation (9). ARIMA models of y_{gt} are the competitors to all these versions. One of the five criteria that Swamy and Tavlas (2001) use to choose among different models is predictive testing – extrapolation to data outside the sample. To facilitate the use of this criterion, we, in this section, give our own answer to the question: How to measure the goodness of forecasts generated from different versions of Equation (9) and their competitors? We will also show in this section how to derive an optimal predictor of the out-of-sample values of the dependent variable of (9).

Let T_g be the ending month of an estimation period and let y_{g,T_g+f} denote an employment relative for small area g in an out-of-sample month $t = T_g + f$. To generate the forecasts of y_{g,T_g+f} from the models of y_{gt}, we may either fix T_g or fix f. By fixing T_g and varying f we obtain multi-step-ahead forecasts and by fixing f and varying T_g we obtain f-step-ahead forecasts. Let $\hat{y}_{g,T_g+f}$ denote a predictor of y_{g,T_g+f} given by a version of Equation (9). Then the values of $\hat{y}_{g,T_g+f}$, f = 1, 2, ..., F, with fixed T_g denote multi-step-ahead forecasts and the values of $\hat{y}_{g,T_g+f}$, $T_g = n+1, n+2, \ldots, n+\tau$, with fixed f denote f-step-ahead forecasts. For ease of exposition, we only consider the multi-step-ahead forecasts below. For such forecasts, the root mean squared error (RMSE) is given by the expression

$$\sqrt{\frac{1}{F}\sum_{f=1}^{F}(\hat{y}_{g,T_g+f} - y_{g,T_g+f})^2}. \tag{10}$$

This has an obvious scaling problem. Several measures that do not are

$$\sqrt{\frac{1}{F}\sum_{f=1}^{F}\left(\frac{\hat{y}_{g,T_g+f} - y_{g,T_g+f}}{y_{g,T_g+f}}\right)^2}, \tag{11}$$

Theil's statistics:

$$U = \sqrt{\frac{(1/F)\sum_{f=1}^{F}(\hat{y}_{g,T_g+f} - y_{g,T_g+f})^2}{(1/F)\sum_{f=1}^{F} y_{g,T_g+f}^2}}, \tag{12}$$

$$U_\Delta = \sqrt{\frac{(1/F)\sum_{f=1}^{F}(\delta^d \delta_s^D \hat{y}_{g,T_g+f} - \delta^D \delta_s^D y_{g,T_g+f})^2}{(1/F)\sum_{f=1}^{F}(\delta^d \delta_s^D y_{g,T_g+f})^2}}, \tag{13}$$

where $\delta^d = (1 - B)^d$ and $\delta_s = (1 - B^s)$ when B is a backward shift operator defined as $By_t = y_{t-1}$ or more generally as $B^s y_t = y_{t-s}$, $(1 - B)^d y_t$ is the first difference of y_t taken d times, and $\delta_s^D y_t = (1 - B^s)^D y_t$ is the difference $y_t - y_{t-s}$ taken D times.

A competitor to the above measures is Relative Mean Absolute Error (RMAE):

$$\frac{1}{F}\sum_{f=1}^{F}\frac{|\hat{y}_{g,T_g+f} - y_{g,T_g+f}|}{y_{g,T_g+f}}, \tag{14}$$

when $y_{g,T_g+f} > 0$. The relative mean squared error (MSE) formula of Montgomery et al. (1998) is different from (11)–(14). For Equation (9), that formula can be expressed as

$$\text{Relative MSE} = \frac{\text{MSE of } f\text{-step-ahead forecasts from Equation (9)}}{\text{MSE of } f\text{-step-ahead forecasts from a benchmark model}}, \tag{15}$$

where f has the same value in both the numerator and the denominator. The limitation of this relative MSE is that it does not reflect the absolute increases in its numerator.

The statistical theory of forecasting shows that formulas (10)–(15) are not always valid. When they are invalid, their use may lead to the choices of bad models. We state below the conditions under which formulas (10)–(15) are valid. The problem we are dealing with here is the determination of a suitable predictor, that is, a real-valued function $\hat{y}_{g,T_g+f}$ defined over a sample space, of which it is hoped that $\hat{y}_{g,T_g+f}$ will tend to be close to the actual value y_{g,T_g+f}. A predictor $\hat{y}_{g,T_g+f}$ is to be close to y_{g,T_g+f} and, since $\hat{y}_{g,T_g+f}$ is a random variable, we shall interpret this to mean that the actual value y_{g,T_g+f} is covered by an interval of values, which $\hat{y}_{g,T_g+f}$ takes with a high probability. To make this requirement precise, we specify three measures of the closeness of (or distance from) a predictor to y_{g,T_g+f}.

Let $\tilde{y}_{g,T_g+f}$ denote a predictor of y_{g,T_g+f} given by a competitor to the version of Equation (9) that gave $\hat{y}_{g,T_g+f}$.

Criterion of Highest Concentration

The predictor $\hat{y}_{g,T_g+f}$ is better than the predictor $\tilde{y}_{g,T_g+f}$ if their distributions satisfy the condition

$$\begin{aligned} &pr(y_{g,T_g+f}-\lambda_1 < \hat{y}_{g,T_g+f} < y_{g,T_g+f} + \lambda_2) \geq \\ &pr(y_{g,T_g+f}-\lambda_1 < \tilde{y}_{g,T_g+f} < y_{g,T_g+f} + \lambda_2) \end{aligned} \tag{16}$$

for all possible values of λ_1 and λ_2 in a chosen interval $(0, \lambda)$ for all possible realizations y_{g,T_g+f} (Rao, 1973, p. 315).

Unfortunately, the predictors that satisfy condition (16) do not exist, as the following theorem shows.

Oakes' (1985) Theorem

There is no universal algorithm to guarantee accurate forecasts forever.

We are not aware of any forecasting work done in the past that contradicts this theorem. More recent evidence that supports the theorem comes from Montgomery et al. (1998). After generating the out-of-sample forecasts from a variety of linear and nonlinear time series models for quarterly and monthly U.S. unemployment rate, Montgomery et al. conclude: "All of these forecasts share certain strengths and weaknesses, but clearly they are not perfectly correlated, and none dominates the others" (p. 488). So we cannot expect one set of regressors, one set of concomitants, and one set of the estimates of the coefficients of Equation (9) to yield accurate forecasts of employment for all future periods, even for a single small area. However, it is possible that one model dominates in a specific region/state compared with others in a finite forecast period.

A *necessary* condition for condition (16) to be satisfied for all λ is that

$$E(\hat{y}_{g,T_g+f}-y_{g,T_g+f})^2 \leq E(\tilde{y}_{g,T_g+f}-y_{g,T_g+f})^2 ; \tag{17}$$

that is, the mean square error of $\hat{y}_{g,T_g+f}$ about the actual value y_{g,T_g+f} is a minimum (Rao, 1973, p. 315). If this condition does not hold, then condition (16) fails. Furthermore, it can be seen from (17) that if $\hat{y}_{g,T_g+f}$ does not possess a finite mean square error, then any predictor with finite mean square error dominates $\hat{y}_{g,T_g+f}$ and $\hat{y}_{g,T_g+f}$ does not satisfy either the criterion of highest concentration in (16) or the minimum mean square error condition in (17). In this case, formulas (10)–(12) and (15) are inappropriate.

Large values of formulas (10)–(12) and (15) indicate a low accuracy of $\hat{y}_{g,T_g+f}$ if the loss attached to the forecast errors, $\hat{y}_{g,T_g+f} - y_{g,T_g+f}$, $f = 1, 2, \ldots, F$, is the sum of (relative) squared errors. Similarly, large values of formula (14) indicate a low accuracy of $\hat{y}_{g,T_g+f}$ if the loss attached to the forecast errors, $\hat{y}_{g,T_g+f} - y_{g,T_g+f}$,

$f = 1, 2, \ldots, F$, is the sum of (relative) absolute errors. These functional forms for the loss may not be appropriate for the following reasons. The adverse social impact of unemployment is higher during contraction periods than during expansion periods. Therefore, the loss attached to the forecast error $\hat{y}_{g,T_g+f} - y_{g,T_g+f}$ is greater if $T_g + f$ occurs during a contraction period than if it occurs during an expansion period (Montgomery et al., 1998, pp. 478, 485). In other words, if $L(\hat{y}_{g,T_g+f}, y_{g,T_g+f})$ represents a loss in predicting y_{g,T_g+f} by $\hat{y}_{g,T_g+f}$, then it is neither (relative) squared error nor (relative) absolute error and the formulas (10)–(15) are inappropriate. However, the squared- or absolute-error loss function may provide a good local approximation to the correct loss function one should use in this situation. If we knew the correct functional form of $L(\hat{y}_{g,T_g+f}, y_{g,T_g+f})$, then we could use the following definition of forecast accuracy.

Pitman's Nearness

The predictor $\hat{y}_{g,T_g+f}$ is nearer to the value y_{g,T_g+f} than the predictor $\tilde{y}_{g,T_g+f}$ if

$$pr[L(\hat{y}_{g,T_g+f}, y_{g,T_g+f}) < L(\tilde{y}_{g,T_g+f}, y_{g,T_g+f})] > \tfrac{1}{2} \tag{18}$$

(Peddada, 1985).

This definition eliminates the following difficulty caused by criteria (16) and (17): The predictor $\hat{y}_{g,T_g+f}$ with no finite mean square error may satisfy condition (18), even though it does not satisfy conditions (16) and (17). Criterion (18) may actually choose such a predictor, which criteria (16) and (17) rule out a priori.

If the value y_{g,T_g+f} is unknown, as it is if $T_g + f$ is a period that has not yet occurred, then the minimum mean square error predictor defined in (17) involves this unknown value and hence is not operational (Lehmann and Casella, 1998, p. 212, Problems 3.6(a) and 3.7(a)). For this reason, optimal predictors are found in the statistics literature by using the criterion:

Minimum Average Mean Square Error

$$\min_{\hat{y}_{g,T_g+f}} E(\hat{y}_{g,T_g+f} - Y_{g,T_g+f})^2, \tag{19}$$

where the expectation is taken with respect to variations in both $\hat{y}_{g,T_g+f}$ and Y_{g,T_g+f}, whose values we have been denoting by y_{g,T_g+f}.

If we impose only minimal restrictions on the form of $\hat{y}_{g,T_g+f}$, then the solution to the minimization problem in (19) is extremely simple. For we know from the Cramer and Doob result (Rao, 1973, p. 264) that when we have the vector of random variables $(Y_{g1}, Y_{g2}, \ldots, Y_{gT_g})'$ in addition to Y_{g,T_g+f}, and wish to predict the value y_{g,T_g+f} from a Borel measurable function $\hat{y}_{g,T_g+f} = f(Y_{g1}, Y_{g2}, \ldots, Y_{gT_g})$, then among all such functions with *finite second-order moments*, that which minimizes the *average* mean square error in (19) is the conditional expectation of

Y_{g,T_g+f}, given $Y_{g1} = y_{g1}$, $Y_{g2} = y_{g2}, \ldots, Y_{gT_g} = y_{gT_g}$, i.e., the lower bound of the average mean square error in (19) is attained when

$$f(y_{g1}, y_{g2}, \ldots, y_{gT_g}) = E(Y_{g,T_g+f}|y_{g1}, y_{g2}, \ldots, y_{gT_g}). \tag{20}$$

The proof of this result can be traced back to Cramer and Doob. That is why we call it the Cramer and Doob result.

Can we determine when the conditional expectation in (20) gives good forecasts of y_{g,T_g+f} and when it gives poor forecasts? An answer is as follows. It should be recognized that the use of the minimum average mean square error predictor (20) requires that attention be restricted to predictors that possess finite second-order moments and that the loss attached to forecast errors be the sum of squared errors. Unfortunately, the sum of squared-error loss function is inappropriate to the problem of forecasting employment numbers, as we have already shown. Another difficulty is that the 'true' functional form of the conditional expectation in (20) is usually unknown. Misspecifications of conditional expectations occur all the time in practice. To quote Montgomery et al. (1998, p. 480), "all models are misspecified ...". Conditional expectations with wrong functional forms may generate poor forecasts. To guard against such possibilities, we have employed in the previous section a rich class of functional forms to derive the conditional expectation of Y_{g,T_g+f} without ignoring omitted-regressors and measurement-error bias. Montgomery et al. (1998) cannot say that Equation (9) is as badly misspecified as models which they say are misspecified. The minimum average mean square error criterion in (19) is useful for our purpose of generating small-area employment forecasts, provided we are willing to ignore the fact that $L(\hat{y}_{g,T_g+f}, y_{g,T_g+f})$ is neither (relative) squared error nor (relative) absolute error. As we have mentioned above, changing the set of concomitants in Equation (7), we can generate from Equation (9) a variety of functional forms for the conditional expectation in (20). Assumptions I and II imply that the average mean square error in (19) is finite. In these cases, Theil's U statistics in (12) is a valid measure of forecast accuracy, provided, of course, $L(\hat{y}_{g,T_g+f}, y_{g,T_g+f})$ is (relative) squared error summed over f.

Now we outline briefly the proof of the Cramer and Doob result. For any predictor, say f, which is a Borel function of $(Y_{g1}, Y_{g2}, \ldots, Y_{gT_g})'$,

$$\begin{aligned} &E[Y_{g,T_g+f} - E(Y_{g,T_g+f}|y_{g1}, y_{g2}, \ldots, y_{gT_g})] \\ &\quad [E(Y_{g,T_g+f}|y_{g1}, y_{g2}, \ldots, y_{gT_g}) - f] = 0 \end{aligned} \tag{21}$$

so that

$$\begin{aligned} E(Y_{g,T_g+f} - f)^2 &= E[Y_{g,T_g+f} - E(Y_{g,T_g+f}|y_{g1}, y_{g2}, \ldots, y_{gT_g})]^2 + \\ &\quad + E[E(Y_{g,T_g+f}|y_{g1}, y_{g2}, \ldots, y_{gT_g}) - f]^2 \\ &\geq E[Y_{g,T_g+f} - E(Y_{g,T_g+f}|y_{g1}, y_{g2}, \ldots, y_{gT_g})]^2, \end{aligned} \tag{22}$$

which is the lower bound for the average mean square error $E(Y_{g,T_g+f} - f)^2$. This lower bound is attained when condition (20) is satisfied, showing that the best choice of the predictor that minimizes the average mean square error in (19) is the conditional expectation in (20). This result critically depends on condition (21) that holds if $E|Y_{g,T_g+f}| < \infty$ and $E|Y_{g,T_g+f}f| < \infty$ (Rao, 1973, p. 97, (2b.3.8)). These conditions are not satisfied when Y_{g,T_g+f} follows a random walk or, more generally, an ARIMA model.

To see this, suppose that Y_{g,T_g+f} follows a random walk model, denoted by $Y_{g,T_g+f} = Y_{g,T_g+f-1} + a_{g,T_g+f}$, where a_{g,T_g+f} is white noise with mean zero and finite variance. Then $Y_{g,T_g+f} = \sum_{j=0}^{\infty} a_{g,T_g+f-j}$. From this equation it follows that Y_{g,T_g+f} does not possess a finite unconditional variance, even though its conditional variance given y_{g,T_g+f-1} is finite, implying that the unconditional average mean square error in (19) is not finite (Rao, 1973, p. 111, (2c.2.6)).[3] No infinite value can be minimized. Thus, the minimization problem in (19) is not solvable if Y_{g,T_g+f} follows a random walk. Extending this result to ARMA models shows that under squared-error loss, the conditional expectation of Y_{g,T_g+f}, given its lagged values implied by an ARIMA model, has no finite unconditional variance and hence is an unconditionally inadmissible predictor of y_{g,T_g+f}. Any predictor of y_{g,T_g+f} with a finite mean square error dominates this predictor. The unconditional expectations of formulas (10)–(12), (14) and (15) are not finite in this case. The conditional expectations of variables following ARIMA models cannot be optimal in the sense of conditions (16) and (17). However, the conditional expectation of $\delta^d \delta_s^D Y_{g,T_g+f}$, given its lagged values implied by an ARIMA$(p, d, q)(P, D, Q)$ model of Y_{g,T_g+f}, possesses a finite unconditional variance and is the minimum average mean square error predictor of $\delta^d \delta_s^D y_{g,T_g+f}$. The minimum average mean square error predictor of y_{g,T_g+f} may not exist when the minimum average mean square error predictor of $\delta^d \delta_s^D y_{g,T_g+f}$ exists. This problem presents itself not only when $\delta^d \delta_s^D y_{g,T_g+f}$ is assumed to follow an ARMA model but also when $\delta^d Y_{g,T_g+f}$ is assumed to follow a co-integrating or an error-correction model described in Greene (2000, pp. 793–796). Similar problems arise in the estimation of regression coefficients. Brown (1990) rigorously proved that a conditionally admissible estimator of regression coefficients could be unconditionally inadmissible. We know that inadmissible predictors are bad, but do not know which one of the conditional- and unconditional-inadmissibility properties is worse.

All this is not to say that the conditional expectations implied by ARIMA models do not give good approximations to unconditionally admissible predictors of the values of their respective dependent variables in some periods. We cannot, however, know a priori when such good approximations will occur. What can be said is that the conditional expectation of $\delta^d \delta_s^D Y_{g,T_g+f}$ implied by an ARIMA(p, d, q) model of Y_{g,T_g+f} with a multiplicative seasonal factor ARIMA(P, D, Q) minimizes the average mean square error $\mathrm{E}(\delta^d \delta_s^D \hat{y}_{g,T_g+f} - \delta^d \delta_s^D Y_{g,T_g+f})^2$ and formula (13) is appropriate if the loss attached to the forecast error $\delta^d \delta_s^D \hat{y}_{g,T_g+f} - \delta^d \delta_s^D y_{g,T_g+f}$ is (relative) squared error. This formula gives a better measure of the accuracy of forecasts from ARIMA$(p, d, q)(P, D, Q)$ models than formulas

(10)–(12), (14) and (15). Because the values of formula (13) are not comparable to those of formulas (10)–(12), (14) and (15), it is inappropriate to reject model (9) in favor of any ARIMA, co-integrating, or error-correction model of a difference of Y_{gt} whenever the value of formula (13) implied by the latter model is smaller than the values of formulas (10)–(12), (14) and (15) implied by the former model.[4] Furthermore, ARIMA models may not have the ability to track turning points in the employment data. For example, Montgomery et al.'s (1998, p. 487) ARIMA(1, 1, 0) model fitted to U.S. quarterly unemployment rate data, being a linear model, lacks the ability to reproduce the countercyclical movements of the unemployment rate. Their seasonal ARIMA(1, 1, 0)(4, 0, 4) model fitted to the same data does allow for complex roots in the autoregressive polynomial, which allow for turning points in the forecasts. The model is still not a good one because it under-predicts the unemployment rate during the rapid increase of 1982 and exhibits forecasts that fluctuate a great deal more during stable periods of unemployment (see Montgomery et al., 1998, p. 487).

It follows from Rao (1973, p. 96, (2b.2.7)) that if F_1 and F_2 are the distribution functions of the predictors $\hat{y}_{g,T_g+f}$ and $\tilde{y}_{g,T_g+f}$ of y_{g,T_g+f} with means μ_1, μ_2 and finite variances σ_1^2, σ_2^2, respectively, then the inequality

$$F_1(y_{g,T_g+f} + \mu_1) - F_1(-y_{g,T_g+f} + \mu_1) \geq F_2(y_{g,T_g+f} + \mu_2) - F_2(-y_{g,T_g+f} + \mu_2) \tag{23}$$

for each y_{g,T_g+f} implies that $\sigma_1^2 \leq \sigma_2^2$. However, the converse of this statement is not true. That is, a *necessary* condition for the inequality in (23) to be satisfied is that the variance of the predictor $\hat{y}_{g,T_g+f}$ is a minimum. This explains why condition (17) is necessary but not sufficient for condition (16) to be true. Unfortunately, even the necessary condition $\sigma_1^2 \leq \sigma_2^2$ cannot be verified unless the population values of σ_1^2 and σ_2^2 are known.[5]

The best linear unbiased predictor (BLUP) of y_{g,T_g+f} can be found by solving the minimization problem in (19) subject to (i) the linearity restriction that the predictor $\hat{y}_{g,T_g+f} = f(Y_{g1}, Y_{g2}, \ldots, Y_{gT_g})$ is a linear homogeneous function, and (ii) the 'unbiasedness' restriction that $E(\hat{y}_{g,T_g+f} - Y_{g,T_g+f}) = 0$. This 'unbiasedness' restriction ensures that the distributions of $\hat{y}_{g,T_g+f}$ and Y_{g,T_g+f} are located at the same value so that their variances are comparable.

There is a difficulty with the BLUP, however. The means μ_1 and μ_2 are equal to EY_{g,T_g+f} if the predictors considered in (23) are unbiased. Thus, if $\hat{y}_{g,T_g+f}$ is the BLUP of y_{g,T_g+f}, then a necessary condition for the probability of $|\hat{y}_{g,T_g+f} - EY_{g,T_g+f}| \leq y_{g,T_g+f}$ to be a maximum is satisfied for all possible realizations y_{g,T_g+f}. This result does us little good if we cannot find the BLUPs of y_{g,T_g+f} that have the same mean value as Y_{g,T_g+f}, whose 'true' distribution is unknown. If we ignore omitted-regressors and measurement-error bias when it is present, then we should be lucky not to have our attempts to find the BLUPs fail. Therefore, Equation (9) does the right thing: it does not ignore such a bias. Yokum, Wildt and

Swamy (1998) use simulation experiments to show that Equation (9) has better predictive properties than its fixed coefficients versions. Conditions (16)–(23) were previously used in Swamy and Schinasi (1989).

The forecasts generated from a specified conditional expectation are not coherent if this conditional expectation is based on contradictory assumptions. An example of a pair of assumptions that contradict each other is: (i) differencing of Y_{g,T_g+f} induces stationarity and (ii) the coefficients of the models of Y_{g,T_g+f} change over time. Taking the first differences of variables on both sides of Equation (6) shows this contradiction. Even an inconsistent model based on contradictory assumptions may sometimes predict the values of its dependent variable very accurately. Therefore, the good predictive performance of a model in some periods might mislead us into thinking that the model is logically valid when in fact it is not. How can we guard ourselves against such a possibility? A good Bayesian, de Finetti, gives an answer. We can never know the 'true' functional form of the conditional expectation in (20) even if such a form exists. We can make assumptions about it. To satisfy a necessary condition under which a specified functional form of (20) is true, de Finetti (1974) sets up minimal coherence criteria that forecasts should satisfy based on data currently available. By these criteria, different forecasts are equally valid *now* if they all satisfy the requirements for coherence, given currently available knowledge. A forecast from an estimated conditional expectation can best represent a measure of the confidence with which one expects that conditional expectation to predict an event in the future, based on currently available data and not on data yet to be observed, provided that the forecast satisfies the requirements for coherence (Schervish, 1985).

Economists have long recognized that no two business cycles that had occurred in the past closely resemble each other in shape. If business cycles that have occurred in different periods have different shapes, then they represent a time-varying environment. In this environment, the coefficients of Equation (6) may not be constant. Equation (6) with changing coefficients represents a time-varying environment. We only consider coherent variations in these coefficients, i.e., variations that are consistent with the real-world interpretations of the coefficients. We do not assume unit-root nonstationarity in a time-varying environment to avoid contradictions. With contradictions, we cannot obtain coherent forecasts. In time-varying environments, the variance of $\hat{y}_{g,T_g+f}$ may change over time and the sum of (relative) squared-error or absolute-error loss in (10)–(15) may be inappropriate. An appropriate formula for assessing the predictive accuracy of Equation (9) in these environments may be

$$\left| \frac{\hat{y}_{g,T_g+f} - y_{g,T_g+f}}{y_{g,T_g+f}} \right| \tag{24}$$

with no summation over the given values of f. By comparing the values of this formula for different models in each month of a forecasting period we may be able to isolate those models that frequently perform well in prediction. Performing frequently well in prediction is a characteristic of good models.

It may not be correct to take the average of (24) over the three months in each quarter of a forecast period because quarterly aggregates of Equation (9) no longer have the same model structure as Equation (9). This leads to model inconsistencies that complicate the modeling process and require further study.

All the discussion given in this section is irrelevant if our objective is to measure the accuracy of the predictions of y^*_{g,T_g+f}, the 'true' value of y_{g,T_g+f}, generated from Equation (9). Formulas (10)–(15) cannot be evaluated if the unobserved y^*_{g,T_g+f} is used in place of y_{g,T_g+f} used in these formulas. What can be said is that Equation (9) is better equipped to give good forecasts of y^*_{g,T_g+f} than any other equation that ignores omitted-variables and measurement-error bias.

4. Consistent Method of Forecasting

In this section, we discuss an iterative method for evaluating the minimum variance linear unbiased estimators and the BLUPs of the parameters and the error terms of Equation (9), respectively. A matrix formulation of (7) is

$$\gamma_{gt} = \bar{\gamma} + Az_{gt} + \mu_g + L\varepsilon_{gt}\,, \tag{25}$$

where γ_{gt} is the $(p+1)$-vector having the i-th coefficient of Equation (6) as its i-th element, $\bar{\gamma}$ is the $(p+1)$-vector having $\bar{\gamma}_i$ as its i-th element, A is the $(p+1) \times (K-1)$ matrix having α_{ih} as its (i, h) element, z_{gt} is the $(K-1)$-vector having z_{hgt} as its h-th element, L is the $(p+1) \times m$ known matrix having $l_{i\ell}$ as its (i, ℓ) element, μ_g and ε_{gt} are as in Assumption 1.

Inserting (25) into $y_{gt} = x'_{gt}\gamma_{gt}$, which is a matrix formulation of Equation (6), gives the following matrix formulation of Equation (9):

$$y_{gt} = x'_{gt}\bar{\gamma} + (z'_{gt} \otimes x'_{gt})\text{vec}(A) + x'_{gt}\mu_g + x'_{gt}L\varepsilon_{gt}\,, \tag{26}$$

where $x_{gt} = (1, y_{g,t-1}, \ldots, y_{g,t-p})'$ is a $(p+1)$-vector, a prime denotes transposition, $\otimes$ denotes a Kronecker product, and vec(A) is a column stack of A, giving a $(p+1)(K-1)$-vector. A body of T_g time-series observations on y_{gt}, x_{gt}, and z_{gt} for small area g is represented as

$$y_g = X_g\bar{\gamma} + X_{zg}\text{vec}(A) + X_g\mu_g + D_{xg}(I_{T_g} \otimes L)\varepsilon_g\,, \tag{27}$$

where $y_g = (y_{g1}, y_{g2}, \ldots, y_{gT_g})'$ is a T_g-vector of observations on the regressand of Equation (6), $X_g = (x_{g1}, x_{g2}, \ldots, x_{gT_g})'$ is a $T_g \times (p+1)$ matrix of observations on the regressors of Equation (6), $X_{zg} = (z_{g1} \otimes x_{g1}, z_{g2} \otimes x_{g2}, \ldots, z_{gT_g} \otimes x_{gT_g})'$ is a $T_g \times (p+1)(K-1)$ matrix of observations on the concomitants and their interactions with the regressors of Equation (6), $D_{xg} = \text{diag}(x'_{g1}, x'_{g2}, \ldots, x'_{gT_g})$ is a $T_g \times T_g(p+1)$ block-diagonal matrix having the rows of X_g as its diagonal blocks and zeroes elsewhere, I_{T_g} is an identity matrix of order T_g, and $\varepsilon_g = (\varepsilon'_{g1}, \varepsilon'_{g2}, \ldots, \varepsilon'_{gT_g})'$ is a $T_g m$-vector of errors.

Under Assumptions 1 and 2, the conditional expectation of y_g, given the x_{gt} and z_{gt}, is $X_g\bar{\gamma} + X_{zg}\text{vec}(A)$ and its conditional variance is

$$\Omega_{gg} = X_g \Delta X'_g + D_{xg}(I_{T_g} \otimes L)\sigma_g^2 \Sigma_{gg\varepsilon}(I_{T_g} \otimes L')D'_{xg}\,, \tag{28}$$

where $\sigma_g^2\Sigma_{gg\varepsilon}$ is the $T_g m \times T_g m$ covariance matrix of ε_g, the t-th diagonal block of $\Sigma_{gg\varepsilon}$ is equal to the $m\times m$ matrix $E(\varepsilon_{gt}\varepsilon'_{gt})/\sigma_g^2 = \Gamma_g/\sigma_g^2 = \varphi_g(\Gamma_g/\sigma_g^2)\varphi'_g + \Delta_g$, and the (t, s) off-diagonal block of $\Sigma_{gg\varepsilon}$ is equal to the $m \times m$ matrix $E(\varepsilon_{gt}\varepsilon'_{gs})/\sigma_g^2 = \varphi_g^{t-s}(\Gamma_g/\sigma_g^2)$ if $t > s$ and is equal to the $m \times m$ matrix $E(\varepsilon_{gt}\varepsilon'_{gs})/\sigma_g^2 = (\Gamma_g/\sigma_g^2)(\varphi'_g)^{s-t}$ if $s > t$.

The G equations in (27) may be written together as

$$y = X_*(\bar{\gamma}', (\text{vec}(A))')' + D_X\mu + D_{xL}\varepsilon, \tag{29}$$

where $y = (y_1, y_2, \ldots, y_G)'$ is a $(\sum_{g=1}^G T_g)$-vector, $X_* = (X'_{*1}, X'_{*2}, \ldots, X'_{*G})'$ is a $\sum_{g=1}^G T_g \times K(p+1)$ matrix with $X_{*g} = (X_g, X_{zg})$, $(\bar{\gamma}', (\text{vec}(A))')'$ is a $K(p+1)$-vector of coefficients, $D_X = \text{diag}(X_1, X_2, \ldots, X_G)$ is a $\sum_{g=1}^G T_g \times G(p+1)$ block-diagonal matrix having the X_g's $(g = 1, 2, \ldots, G)$ as its diagonal blocks and zeroes elsewhere, $\mu = (\mu'_1, \mu'_2, \ldots, \mu'_G)'$ is a $G(p+1)$-vector of small-area effects, $D_{xL} = \text{diag}(D_{x1}(I_{T_1} \otimes L), D_{x2}(I_{T_2} \otimes L), \ldots, D_{xG}(I_{T_G} \otimes L))$ is a $\sum_{g=1}^G T_g \times \sum_{g=1}^G mT_g$ block-diagonal matrix having the $D_{xg}(I_{T_g} \otimes L)$ $(g = 1, 2, \ldots, G)$ as its diagonal blocks and zeroes elsewhere, and $\varepsilon = (\varepsilon'_1, \varepsilon'_2, \ldots, \varepsilon'_G)'$ is a $(\sum_{g=1}^G mT_g)-$ vector.

The minimum variance linear unbiased estimator of $(\bar{\gamma}', (\text{vec}(A))')'$ in Equation (29) is

$$(\hat{\bar{\gamma}}'(\text{vec}(\hat{A}))')' = (X'_*\Omega^{-1}X_*)^{-1}X'_*\Omega^{-1}y, \tag{30}$$

where $\Omega = \text{diag}(\Omega_{11}, \Omega_{22}, \ldots, \Omega_{GG})$ is a $\sum_{g=1}^G T_g \times \sum_{g=1}^G T_g$ block-diagonal matrix having the Ω_{gg}'s $(g = 1, 2, \ldots, G)$ in (28) as its diagonal blocks and zeroes elsewhere. The covariance matrix of estimator (30) is $(X'_*\Omega^{-1}X_*)^{-1}$.

The means of μ and ε in Equation (29) are zero and their covariance matrices are $(I_G \otimes \Delta)$ and $\Sigma_\varepsilon = \text{diag}(\sigma_1^2\Sigma_{11\varepsilon}, \sigma_2^2\Sigma_{22\varepsilon}, \ldots, \sigma_G^2\Sigma_{GG\varepsilon})$, which is a $\sum_{g=1}^G T_g \times \sum_{g=1}^G T_g$ block-diagonal matrix having the $\sigma_g^2\Sigma_{gg\varepsilon}$'s $(g = 1, 2, \ldots, G)$ in Equation (28) as its diagonal blocks and zeroes elsewhere, respectively. Let $M = [I - X_*(X'_*\Omega^{-1}X_*)^{-1}X'_*\Omega^{-1}]$.

Then the BLUP of μ is

$$\begin{aligned}\hat{\mu} &= (I_G \otimes \Delta)D'_X M'[M(D_X(I_G \otimes \Delta)D'_X + D_{xL}\Sigma_\varepsilon D'_{xL})M']^- My \\ &= (I_G \otimes \Delta)D'_X(D_X(I_G \otimes \Delta)D'_X + D_{xL}\Sigma_\varepsilon D'_{xL})^{-1}My,\end{aligned} \tag{31}$$

where the symbol $[\cdot]^-$ denotes a generalized inverse of the matrix in square brackets, and its covariance matrix is $(I_G \otimes \Delta)D'_X\Omega^{-1}MD_X(I_G \otimes \Delta)$. The BLUP of ε is

$$\begin{aligned}\hat{\varepsilon} &= \Sigma_\varepsilon D'_{xL}M'[M(D_X(I_G \otimes \Delta)D'_X + D_{xL}\Sigma_\varepsilon D'_{xL})M']^- My \\ &= \Sigma_\varepsilon D'_{xL}(D_X(I_G \otimes \Delta)D'_X + D_{xL}\Sigma_\varepsilon D'_{xL})^{-1}My,\end{aligned} \tag{32}$$

and its covariance matrix is $\Sigma_\varepsilon D'_{xL}\Omega^{-1}MD_{xL}\Sigma_\varepsilon$. The derivation of formulas (31) and (32) follows that of Swamy and Mehta (1975) and Chang, Hallahan and Swamy (1992). We can use the estimator in (30) and the predictors in (31) and (32) to obtain the predictor

$$\hat{\gamma}_{gt} = \hat{\bar{\gamma}} + \hat{A}z_{gt} + \hat{\mu}_g + L\hat{\varepsilon}_{gt}, \tag{33}$$

for all g and t.

Estimator (30) and predictors (31) and (32) involve the unknown parameters, Δ, φ_g, Δ_g, and σ_g^2, g = 1, 2,..., G. A method of estimating the parameters φ_g, Δ_g, and σ_g^2 for all g is given in Chang, Swamy, Hallahan and Tavlas (2000). To estimate Δ consider the generalized least squares estimator of $(\bar{\gamma} + \mu_g)$ in Equation (27). This estimator can be written as

$$b_g = J'(X'_{*g}\Sigma_{ggy}^{-1}X_{*g})^{-1}X'_{*g}\Sigma_{ggy}^{-1}y_g, \tag{34}$$

where $J' = [I\ 0]$ is a $(p+1)\times K(p+1)$ matrix having the columns of an identity matrix of order $(p+1)$ as its first $(p+1)$ columns and zeroes elsewhere and $\Sigma_{ggy} = D_{xg}(I_{T_g}\otimes L)\sigma_g^2\Sigma_{gg\varepsilon}(I_{T_g}\otimes L')D'_{xg}$. Let

$$S_b = \sum_{g=1}^{G} b_g b'_g - \frac{1}{G}\sum_{g=1}^{G} b_g \sum_{g=1}^{G} b'_g. \tag{35}$$

Then an unbiased estimator of Δ is

$$\tilde{\Delta} = \frac{S_b}{G-1} - \frac{1}{G}\sum_{g=1}^{G}\sigma_g^2 J'(X'_{*g}\Sigma_{ggy}^{-1}X_{*g})^{-1}J. \tag{36}$$

The covariance matrix Δ is a nonnegative definite matrix; hence estimator (36) taking on negative definite values with positive probability is an inadmissible estimator of Δ against any loss function for which the risk function exists. A nonnegative definite estimator of Δ is

$$\hat{\Delta} = \sum_{j=1}^{p_1}\lambda_j q_j q'_j, \tag{37}$$

where the λ_j's $(j = 1, 2, \ldots, p_1 \le (p+1))$ are the nonnegative eigenvalues of $\tilde{\Delta}$ and the q_j's are the corresponding eigenvectors. Rao (1973, p. 63, 1f.2.11) proved that if estimator (37) is a matrix of rank r $(\le p_1)$, then it provides the closest fit to $\tilde{\Delta}$ among all matrices of rank r (see also Chang et al., 2000, p. 127).

Formulas (30)–(37) can be evaluated for given Δ, φ_g, Δ_g, and σ_g^2, $g = 1, 2, \ldots, G$. The estimators of Δ, φ_g, Δ_g, and σ_g^2, $g = 1, 2, \ldots, G$, can be evaluated for given $\bar{\gamma}$ and vec(A). Starting with some initial values for φ_g, Δ_g, and σ_g^2, $g = 1, 2, \ldots, G$, we estimate all the unknown parameters of model

(9) using an iterative re-weighted generalized least squares (IRWGLS) method. In this method, the estimated covariance matrix of the composite error term of Equation (9) changes from iteration to iteration. Hence, the term 're-weighted' appears in IRWGLS. This iteration is continued until convergence. We use Paige's (Kourouklis and Paige, 1981) numerically stable algorithm to evaluate formulas (30)–(37). Suppose that the coefficients and the error terms of Equation (29) are estimated using all observations through $t = T_g$. Then the f-step-ahead forecast of y_{g,T_g+f} implied by these estimates is

$$\begin{aligned}\hat{y}_{g,T_g+f} &= \hat{x}'_{g,T_g+f}(\hat{\bar{\gamma}} + \hat{\mu}_g) + (\hat{z}'_{g,T_g+f} \otimes \hat{x}'_{g,T_g+f})\text{vec}(\hat{A}) + \\ &+ \hat{x}'_{g,T_g+f} L\hat{\varphi}_g^f \hat{\varepsilon}_{g,T_g},\end{aligned} \tag{38}$$

which is obtained from Equation (26) by setting (i) $t = T_g + f$, (ii) $\bar{\gamma}$, $\text{vec}(A)$,φ_g, μ_g, and ε_{g,T_g} equal to their respective sample estimates, and (iii) y_{g,T_g+f-j} with $f > j$ in x_{g,T_g+f} and z_{g,T_g+f} equal to its forecast. This forecast is obtained from Equation (38) by setting $t = T_g + f - j > T_g$. Only the one-step-ahead forecasts of y_{g,T_g+f} with $f = 1$ do not depend on the forecasts of y_{g,T_g+f-j} with $f > j$. Note that $\hat{\mu}_g$ and $\hat{\varepsilon}_{g,T_g}$ are the g-th $(p + 1)$-sub-vector of (31) and the gT_g-th m-sub-vector of (32), respectively. Equation (38) with $\hat{x}_{g,T_g+f}$ and $\hat{z}_{g,T_g+f}$ reflects the actual forecasting environment within which practitioners must operate. Because of the necessity to use the forecasts of y_{g,T_g+f-j} with $f > j$ in this situation it is difficult to find appropriate concomitants that satisfy Assumptions 1 and 2, as we show in the next section.

We need to find the variance of $\hat{y}_{g,T_g+f}$ in (38). An approximation to this variance is provided by

$$\begin{aligned}&[(1, \hat{z}'_{g,T_g+f}) \otimes \hat{x}'_{g,T_g+f}](X'_*\hat{\Omega}^{-1}X_*)^{-1}[(1, \hat{z}'_{g,T_g+f}) \otimes \hat{x}'_{g,T_g+f}]' + \\ &+ \hat{x}'_{g,T_g+f} L\hat{\varphi}_g^f \text{Cov}(\hat{\varepsilon}_{g,T_g})\hat{\varphi}_g^{f'} L'\hat{x}_{g,T_g+f},\end{aligned} \tag{39}$$

where the sample estimates of Δ,φ_g, Δ_g, and σ_g^2 for all g are used in place of their true values used in (30),

$$\begin{aligned}&\text{Cov}(\hat{\varepsilon}_{g,T_g}) = \\ &\quad (\hat{\sigma}_g^2\hat{\Sigma}_{gg\varepsilon})_{m\times T_g m}(I_{T_g} \otimes L')D'_{xg}(\hat{\Omega}^{-1}\hat{M})_{T_g\times T_g} D_{xg}(I_{T_g} \otimes L)(\hat{\sigma}_g^2\hat{\Sigma}_{gg\varepsilon})'_{T_g m\times m}\end{aligned}$$

is a sample estimate of the covariance matrix of $\hat{\varepsilon}_{g,T_g}$, $(\hat{\sigma}_g^2\hat{\Sigma}_{gg\varepsilon})_{m\times T_g m}$ is a sample estimate of the last m rows of $\sigma_g^2\Sigma_{gg\varepsilon}$, and $(\hat{\Omega}^{-1}\hat{M})_{T_g\times T_g}$ is a sample estimate of the g-th diagonal block of $\Omega^{-1}M$.

5. An Empirical Example

In this section, we explain how to generate 1- to 12-month-ahead forecasts of certain out-of-sample values of the dependent variable of Equation (9) using the following values: (i) The values 3, 5, and 6 for p, (ii) the value I for L, (iii) the zero values for the off-diagonal elements of φ_g, (iv) the zero value for Δ, (v) the zero values or the nonzero IRWGLS estimates for the diagonal elements of φ_g, (vi) the corresponding IRWGLS estimates of $\bar{\gamma}$, $A, \Delta_g, \sigma_g^2, \mu_g$, and ε_{g,T_g+f}, and (vii) several sets of concomitants are used in Equation (38) to obtain the 1- to 12-month-ahead forecasts of employment relatives for the forecast period, 1996.04–2001.12, and for 13 small domains.

The nonlinearities of (2) as well as the omitted-variable and measurement-error biases contained in the coefficients of (6) depend on p. Hence, the concomitants that explain the variation in the coefficients of Equation (6) cannot be independent of p. This shows that we need to specify p before we pick the concomitants to be included in Equation (7). Our initial set of experiments show that Equation (9) with $p = 5$ and no zs performs better in prediction than the same equation with $p = 3$ or 6. It performs better in prediction with variable coefficients than with fixed coefficients. The assumption of nonzero serial correlation in (8) improves the forecasting performance of Equation (6). Quite possibly, this result would not obtain if Equation (6) included an adequate number of appropriate concomitants.

For $p = 5$, the concomitants we used include $z_{1gt} = \log(y_{g,t-2}/y_{g,t-3})$, $z_{2gt} = (1/5)\log(y_{g,t-1}/y_{g,t-6})$, $z_{3gt} = (1/5)\log(y_{g,t-7}/y_{g,t-12})$, $z_{4gt} = (y_{g,t-2} - y_{g,t-3})/y_{g,t-3}$, $z_{5gt} = (y_{g,t-1} - y_{g,t-6})/5y_{g,t-6}$, $z_{6gt} = (y_{g,t-7} - y_{g,t-12})/5y_{g,t-12}$, $z_{7gt} = (y_{g,t-1} - y_{g,t-6})/5(y_{g,t-1} + y_{g,t-6})2^{-1}$, $z_{8gt} = (y_{g,t-7} - y_{g,t-12})/5(y_{g,t-7} + y_{g,t-12})2^{-1}$, $z_{9gt} = (y_{g,t-2} - y_{g,t-6})/4(y_{g,t-2} + y_{g,t-6})2^{-1}$, $z_{10gt} = (y_{g,t-3} - y_{g,t-6})/3(y_{g,t-3} + y_{g,t-6})2^{-1}$, and twelve monthly dummies, each of which takes a value of 1 in a specific month and zeroes in other months.

The source of our data is the ES-202 program, which provides monthly data for all small domains in the U.S. for the period January 1990–March 2001. To make use of RMSEs (formula (10)) and Theil's U statistic (formula (12)), we assume that these data do not contain measurement errors. We are not too happy about this assumption because it can be false. Unfortunately, without the assumption, formulas (10)–(14) are not computable. We divide the employment total for each month by its March 1999 value to obtain the dependent variable of Equation (9). The monthly series on this dependent variable for each of several small domains is plotted for January 1990–March 2001. These plots show the following: (i) The employment totals for each of these small domains fluctuate around a nonlinear trend. (ii) Several of these series display pronounced seasonal patterns. (iii) Both the seasonal pattern and the shape of the nonlinear trend are different for different small domains and change over time. (iv) Changes in employment totals in some months are much bigger than changes in other months; these big changes do not occur in the same month for all these series. (v) Mostly, employment totals decrease at a faster rate than they increase. (vi) During the period January 1990–March 2001, all the small domains we examined recurrently experienced periods of increase and decrease in their employment totals, although the length and depth of those cycles were irregular.

In view of Oakes' theorem, we do not expect any one set of the estimates of the coefficients and the error terms of a version of Equation (9) to produce accurate forecasts of y_{g,T_g+f} for all f if T_g is fixed. Consequently, the forecasts from Equation (9) are generated using the rolling forecast method. First, the parameters and the error terms of several versions of Equation (9) are estimated using all observations through a given forecasting origin (T_g); next, 1- to 12-month-ahead forecasts of y_{g,T_g+f}, $f = 1, 2, \ldots, 12$, are generated for this origin. This procedure is then repeated for all the chosen forecasting origins. To reduce our computational burden, we have chosen only the last month of each quarter in the period, January 1996–December 2000, as a forecasting origin. The ending month of an estimation period is advanced one quarter to obtain the next estimation period. With this method, we have 20 estimation periods, the first of which is January 1992–March 1996. For this period, the forecasting origin is March 1996. That is, the ending month of each estimation period is a forecasting origin. In this forecast method, we use the forecasts in place of the actual values of regressors used on the right-hand sides of Equations (6) and (7) for the forecast periods. For example, y_{g,T_g+f-j} with $f > j$ belongs to a forecast period and its forecast is used in place of its actual value used on the right-hand sides of Equations (6) and (7) for $t = T_g + f$. This has led to some disconcerting results, as we show below.

With 13 small domains and 20 estimations for each domain, we have 260 cases. For each of these domains, the 20 estimations gave us 20 sets of 1- to 12-month-ahead forecasts. But we did not have the actual values that were necessary to compute the RMSEs and Theil's U statistic for all these forecasts. Since our data series ended at March 2001, we could only compute Theil's U statistic for the 10- to 12-month-ahead forecasts and the RMSEs for the 1- to 12-month-ahead forecasts obtained in the first 17 estimations for each domain. These computations gave us 221 values of Theil's U statistic and the same number of the values of RMSEs. For the period January 2001–March 2001, we could compute Theil's U statistic for 7- to 9-, 4- to 6-, and 1- to 3-month-ahead forecasts obtained in the 18th, 19th and 20th estimations, respectively. The results of these computations are summarized in Tables I and II. We also computed the RMSEs for 1- to 9-, 1- to 6-, and 1- to 3-month-ahead forecasts obtained in the 18th, 19th, and 20th estimations, respectively.

Our empirical work shows the following:

1. When $p = 5$, the use of $z_{1gt} = \log(y_{g,t-2}/y_{g,t-3})$ or $z_{4gt} = (y_{g,t-2} - y_{g,t-3})/y_{g,t-3}$ as a concomitant led to a high degree of multicollinearity, which slowed down considerably convergence of the IRWGLS method. The same problem arose when $p = 5$ and $z_{9gt} = (y_{g,t-2} - y_{g,t-6})/4(y_{g,t-2} + y_{g,t-6})2^{-1}$ and $z_{10gt} = (y_{g,t-3} - y_{g,t-6})/3(y_{g,t-3} + y_{g,t-6})2^{-1}$ were included in Equation (9) as its concomitants. This high degree of multicollinearity was reduced when we used $z_{2gt} = (1/5)\log(y_{g,t-1}/y_{g,t-6})$, $z_{3gt} = (1/5)\log(y_{g,t-7}/y_{g,t-12})$, $z_{5gt} = (y_{g,t-1} - y_{g,t-6})/5y_{g,t-6}$, $z_{6gt} = (y_{g,t-7} - y_{g,t-12})/5y_{g,t-12}$, $z_{7gt} =$

Table I. Accuracy measures of forecasts constructed from Equation (9) with $p = 5$, no concomitants, and unknown values of y_{g,T_g+f-j} with $f > j$.

SD[a]	Theil's U statistic of forecasts						RMSE of forecasts	
	FP:[b] January 1997–March 2001		FP: January 2001–March 2001				FP: April 1996–March 2001	
	10- to 12-month-ahead		10- to 12-month-ahead	7- to 9-month-ahead	4- to 6-month-ahead	1- to 3-month-ahead	1- to 12-month-ahead	
	Min[c]	Max[c]					Min[d]	Max[d]
1	0.011	0.154	0.090	0.077	0.065	0.023	0.040	0.126
2	0.005	0.165	0.090	0.082	0.064	0.026	0.024	0.135
3	0.013	0.125	0.062	0.015	0.011	0.017	0.012	0.086
4	0.008	0.119	0.045	0.061	0.075	0.048	0.014	0.103
5	0.004	0.143	0.012	0.029	0.039	0.022	0.015	0.119
6	0.007	0.057	0.020	0.008	0.017	0.041	0.024	0.060
7	0.012	0.080	0.028	0.013	0.014	0.054	0.026	0.064
8	0.013	0.210	0.046	0.068	0.023	0.021	0.021	0.140
9	0.004	0.080	0.022	0.014	0.013	0.013	0.012	0.066
10	0.016	0.348	0.016	0.028	0.087	0.014	0.030	0.340
11	0.014	0.314	0.089	0.070	0.052	0.022	0.009	0.278
12	0.007	0.053	0.015	0.029	0.017	0.045	0.019	0.048
13	0.008	0.072	0.033	0.028	0.006	0.017	0.012	0.054

[a] SD = Small Domain; [b] FP = Forecast Period; [c] Min (or Max) = The minimum (or maximum) of 17 values, one for each quarter in January 1997–March 2001; [d] Min (or Max) = The minimum (or maximum) of 17 values, one for each 4 consecutive quarters in April 1996–March 2001.

$(y_{g,t-1}-y_{g,t-6})/5(y_{g,t-1}+y_{g,t-6})2^{-1}$, or $z_{8gt} = (y_{g,t-7}-y_{g,t-12})/5(y_{g,t-7}+y_{g,t-12})2^{-1}$ as a concomitant. Another difficulty is that the variable $\hat{z}_{1gt} = \log(\hat{y}_{g,t-2}/\hat{y}_{g,t-3})$ is not defined when $\hat{y}_{g,t-2}$ or $\hat{y}_{g,t-3}$ is negative, which is the wrong sign. With a high degree of multicollinearity, a negative value for the dependent variable of Equation (38) is not a rarity. A similar problem arose, though less frequently, with Equation (38) when $z_{2gt} = (1/5)\log(y_{g,t-1}/y_{g,t-6})$ or $z_{3gt} = (1/5)\log(y_{g,t-7}/y_{g,t-12})$ appeared as its concomitant. There is no claim that any one set of these concomitants leads to accurate forecasts in all out-of-sample periods.

2. Equation (9) with monthly dummies as its concomitants generated forecasts that fluctuated a great deal more than the actual series.
3. The forecasting performance of Equation (9) deteriorates, especially for the multi-step-ahead forecasts, if an inappropriate concomitant is included in it. (Recall that we consider a concomitant as appropriate if it can satisfy Assumption 2 and explain at least partially the 'true' variation in the coefficients of

Table II. Accuracy measures of forecasts constructed from Equation (9) with $p = 5$, $z_{7gt} = 2(y_{g,t-1} - y_{g,t-6})/5(y_{g,t-1} + y_{g,t-6})$ and $z_{8gt} = 2(y_{g,t-7} - y_{g,t-12})/5(y_{g,t-7} + y_{g,t-12})$ as its concomitants, and with unknown values of y_{g,T_g+f-j} with $f > j$.

SD [a]	Theil's U statistic of forecasts						RMSE of forecasts	
	FP: [b] January 1997–March 2001		FP: January 2001–March 2001				FP: April 1996–March 2001	
	10- to 12-month-ahead		10- to 12-month-ahead	7- to 9-month-ahead	4- to 6-month-ahead	1- to 3-month-ahead	1- to 12-month-ahead	
	Min [c]	Max [c]					Min [d]	Max [d]
1	0.013	0.152	0.130	0.127	0.118	0.047	0.040	0.123
2	0.018	0.098	0.018	0.041	0.026	0.014	0.034	0.114
3	0.007	0.128	0.061	0.018	0.005	0.015	0.012	0.085
4	0.015	0.146	0.034	0.040	0.048	0.024	0.023	0.089
5	0.005	0.116	0.020	0.024	0.008	0.026	0.015	0.173
6	0.007	327.37	0.022	0.011	0.004	0.011	0.010	172.77
7	0.011	56464.	0.011	0.034	0.021	0.034	0.021	26531.
8	0.008	0.248	0.071	0.060	0.028	0.034	0.022	0.162
9	0.005	0.063	0.033	0.012	0.031	0.048	0.011	0.062
10	0.013	0.388	0.026	0.054	0.089	0.012	0.025	0.385
11	0.015	0.69132E+11	0.017	0.019	0.032	0.004	0.009	0.34494E+11
12	0.005	0.053	0.005	0.008	0.010	0.017	0.015	0.041
13	0.014	0.103	0.025	0.032	0.023	0.024	0.018	0.094

[a] SD = Small Domain; [b] FP = Forecast Period; [c] Min (or Max) = The minimum (or maximum) of 17 values, one for each quarter in January 1997–March 2001; [d] Min (or Max) = The minimum (or maximum) of 17 values, one for each 4 consecutive quarters in April 1996–March 2001.

Equation (6).) Even if a variable included in Equation (9) is an appropriate concomitant in this sense, using a very inaccurate value for it for a forecast period leads to a very inaccurate forecast. The values of z implied by the forecasts of y_{g,T_g+f-j} with $f > j$ can be very inaccurate. It is difficult to find the exact values of appropriate concomitants for a future period if these values are unknown. Equation (9) without the z yields forecasts that are not affected by either appropriate or inappropriate concomitants. The accuracy of 10- to 12-month-ahead forecasts from Equation (9) with $p = 5$, no zs, and with the actual values of y_{g,T_g+f-j} with $f > j$ equal to their respective forecasts, as measured by Theil's U statistic, lies between 0.004 and 0.348 for all the 221 cases (Table I). It has increased in 115 of these cases and decreased in the remaining 106 cases when the two variables $z_{7gt} = (y_{g,t-1} - y_{g,t-6})/5(y_{g,t-1} + y_{g,t-6})2^{-1}$ and $z_{8gt} = (y_{g,t-7} - y_{g,t-12})/5(y_{g,t-7} + y_{g,t-12})2^{-1}$ are included as concomitants in Equation (9) with $p = 5$. (Remember that in these cases, we only used the forecasts of y_{g,T_g+f-j} with $f > j$ in place of their actual values used in Equation (6), z_{7gt}, and z_{8gt} for $t = T_g + f$.) That is, in 52% of the 221 cases we considered, Equation (9) with $p = 5$ using the concomitants z_{7gt} and z_{8gt} resulted in an improvement in the predictive ability over Equation (9) with $p = 5$ and no zs when the values of y_{g,T_g+f-j} with $f > j$ appearing in Equation (6), z_{7gt}, and z_{8gt} for $t = T_g + f$ were unknown. While many of the 115 increases in forecast accuracy from using the concomitants z_{7gt} and z_{8gt} are high, only 7 of the 106 decreases in forecast accuracy from using those concomitants are disconcerting. We discuss each of these 7 cases in detail below.

4. Of the 17 values of Theil's U statistic that lie between 0.007 and 327.37 for SD 6 in Table II, 14 are small lying between 0.007 and 0.091 and only three are big being equal to 3.441, 154.66, and 327.37, respectively. These three big values are observed for the forecast periods, January 1997–March 1997, January 2001–March 2000, and October 1998–December 1998, respectively. They get reduced to 0.038, 0.011, and 0.019, respectively, if Equation (9) with $p = 5$, no zs, and with the actual values of y_{g,T_g+f-j} with $f > j$ equal to their forecasts is used to forecast its dependent variable for these forecast periods. The values of z_{7gt} and z_{8gt}, which the forecasts of y_{g,T_g+f-j} with $f > j$ imply for the forecast periods, January 1997–March 1997, October 1998–December 1998, and January 2000–March 2000, are inappropriate.
5. For SD 7 in Table II, only one value of Theil's U statistic for the forecast period January 1997–March 1997 is as high as 56464 and its 16 values for other forecast periods lie between 0.011 and 0.118. This high value can be reduced to 0.023 by not using z_{8gt} as a concomitant. The value of z_{8gt} implied by the forecasts of y_{g,T_g+f-j} with $f > j$ for the forecast period January 1997–March 1997 is inappropriate.
6. The three high values, 0.54437E+08, 0.69132E+11, and 0.9174, are observed for Theil's U statistic in the consecutive forecast periods, October 1998–December 1998, January 1999–March 1999, and April 1999–June 1999, respectively, for SD 11 in Table II. They can be reduced to 0.314, 0.111, and 0.107, respectively, if Equation (9) with $p = 5$, no zs, and with the actual values of y_{g,T_g+f-j} with $f > j$ equal to their forecasts is used to predict its dependent variable for these forecast periods. The values of z_{7gt} and z_{8gt} implied by the forecasts of y_{g,T_g+f-j} with $f > j$ for the forecast periods, October 1998–December 1998, January 1999–March 1999, and April 1999–June 1999, are inappropriate.
7. In several cases we examined, the accuracy of the forecasts from Equation (9) with $p = 5$ and with the two concomitants, z_{7gt} and z_{8gt}, or without the zs does not monotonically decrease as the forecasting horizon is increased (columns 4–7 of Tables I and II). That is, these forecasts can have the desirable property in that the accuracy of 10- to 12-month-ahead forecasts is higher than that of 1- to 9-month-ahead forecasts (lines 4, 5, and 12 of Table I and lines 7

and 12 of Table II). Only the forecasts from nonlinear models can have this property. Equation (9) is a nonlinear model. It can be seen that the errors in the forecasts from a straight line increase in magnitude as the forecasting horizon is increased when the true curve moves away from the straight line in the forecast period. Montgomery et al.'s (1998, p. 484) study of the forecasts from an ARIMA model fitted to quarterly U.S. unemployment rate data shows that there is on average a twenty-fold increase in the MSE of f-step-ahead forecasts as f increases from 1 to 5.

In summary, Equation (9) for $t = T_g + f$ with $p = 5$, no zs, and with the actual values of y_{g,T_g+f-j} with $f > j$ equal to their forecasts yields 10- to 12-month-ahead forecasts of its dependent variable that do not imply a value of Theil's U statistic greater than 0.348 in any of the 221 cases we examined. The accuracy of these forecasts is increased in 52% of the cases when the equation is expanded to include the values of the concomitants, $z_{7gt} = (y_{g,t-1} - y_{g,t-6})/5(y_{g,t-1} + y_{g,t-6})2^{-1}$ and $z_{8gt} = (y_{g,t-7} - y_{g,t-12})/5(y_{g,t-7} + y_{g,t-12})2^{-1}$, implied by the actual values of y_{g,T_g+f-j} with $f \leq j$ and the forecasts of y_{g,T_g+f-j} with $f > j$. In 7 of the 221 cases, the errors contained in the 10- to 12-month-ahead forecasts of the dependent variable of the expanded equation are huge for certain forecast periods. These cases arise not because the expanded equation is not a good nonlinear model but because the forecasts of z_{7g,T_g+f-j}, z_{8g,T_g+f-j}, and y_{g,T_g+f-j} with $f > j$ used are the inappropriate values of z_{7g,T_g+f-j}, z_{8g,T_g+f-j}, and y_{g,T_g+f-j} with $f > j$, respectively. It remains to be determined whether Equation (9) with $p = 5$ and the concomitants, z_{7gt} and z_{8gt}, is the best nonlinear model.

6. Conclusions

The biasing effects of measurement errors, omitted regressors, and of misspecifications of 'true' functional forms are a pervasive problem in econometrics.

The coefficient on a regressor in an equation can only be free of any bias when no relevant regressor is omitted from the equation, when included regressors of the equation are not measured with error, and when the true functional form of the equation is known. This paper explains how the bias-free coefficients on included regressors might be estimated. A method of assessing the predictive accuracy of specified equations is also provided. We have illustrated our methods with an application to the important practical problem of forecasting employment for small domains with nonstationary autoregressive models.

It is true that the forecasts generated from any one model cannot always dominate those from its competitors. A nonlinear autoregressive model that is different from ARIMA models has been developed in this paper to obtain the forecasts of as yet unobserved values of small-area employment. The dependent variable of this nonlinear model has both finite conditional and unconditional variances, while it does not possess a finite unconditional variance, though its conditional variance may be finite, if it follows an ARIMA model instead of the nonlinear model. If the optimality of the predictors of the values of a variable is defined in terms of the smallest of the average unconditional mean square errors of the predictors, then the conditional expectation of the variable implied by an ARIMA model has no finite unconditional variance and hence is unconditionally inadmissible. In time-varying environments, variables do not become stationary after being differenced certain number of times and hence unit-root nonstationarities of ARIMA models might be hard to justify. Our nonlinear autoregressive model is appropriate to time-varying environments. It captures the nonlinearities and nonstationarities inherent in many small-area employment series.

Acknowledgement

The authors are very grateful to David Belsley for making many helpful corrections to their presentation.

Notes

[1] The rationale for using ARIMA models in a time-varying environment of the type represented by Equation (1) is ordinarily very weak because in this environment, differencing does not induce stationarity and hence unit-root nonstationarities of ARIMA models are inappropriate.

[2] With an abuse of notation, we use the same symbol to denote both a random variable and its values at many places in this paper.

[3] The object of filtering is to update our knowledge of a system each time a new observation is brought in. The joint density $p(y_1, \dots, y_T)$ of a set of T variables, $Y_1, \dots, Y_T$, if it exists, can be written as the product of marginal and conditional densities, $p(y_1)p(y_2|y_1)p(y_3|y_2, y_1) \cdots p(y_T|y_{T-1}, \dots, y_1)$. Applications of the Kalman filter to this joint distribution require assumptions about $p(y_1)$. Suppose that Y_t follows a random walk for all t. Then both the joint distribution with density $p(y_1, \dots, y_T)$ and the distribution with density equal to the product, $p(y_1)p(y_2|y_1)p(y_3|y_2, y_1) \cdots p(y_T|y_{T-1}, \dots, y_1)$, have no finite moments, even though the conditional distributions with densities $p(y_2|y_1)$, $p(y_3|y_2, y_1), \cdots, p(y_T|y_{T-1}, \dots, y_1)$, respectively, can have finite moments. This result is contradicted if it is assumed that the unconditional distribution of Y_1 has finite moments. We have seen several Kalman-filter applications that assume that Y_t follows a random walk for all t and the unconditional distribution of Y_1 has finite moments. We have also not seen a single application of the Kalman-filter algorithm that does not ignore the correlations between the regressors of Equation (6) and its coefficients and between γ_{0gt} and the other coefficients of Equation (6).

[4] This point was made previously in Friedman and Schwartz (1991, p. 46).

[5] For example, when Hendry and Ericsson followed the rule that a model variance-dominates another if the *estimated residual* variance of the former is smaller than that of the latter, Friedman and Schwartz (1991, p. 46) pointed out that “they judge the validity of their hypotheses by the data from which they derive them!”

References

Basmann, Robert L. (1988). Causality tests and observationally equivalent representations of econometric models. *J. Econometrics*, **39**, 69–104.

Brown, Laurence D. (1990). An ancillarity paradox which appears in multiple linear regression (with discussion). *Ann. Stat.*, **18**, 471–538.

Chang, I-Lok, Swamy, P.A.V.B., Hallahan, Charles and Tavlas, George S. (2000). A computational approach to finding causal economic laws. *Computat. Econom.*, **16**, 105–136.

Chang, I-Lok, Hallahan, Charles and Swamy, P.A.V.B. (1992). Efficient computation of stochastic coefficient models. In H.M. Amman, D.A. Belsley and L.F. Pau (eds.), *Computational Economics and Econometrics*, Kluwer Academic Publishers, Boston, pp. 43–53.

de Finetti, Bruno (1974). *The Theory of Probability*, Vol. 1. Wiley, New York.

Friedman, Milton and Schwartz, Anna J. (1991). Alternative approaches to analyzing economic data. *Amer. Econom. Rev.*, **81**, 39–49.

Greene, William H. (2000). *Econometric Analysis*, 4th edn. Prentice Hall, Upper Saddle River, New Jersey.

Kourouklis, S. and Paige, C.C. (1981). A constrained least squares approach to the general Gauss-Markov linear model. *J. Amer. Stat. Assoc.*, **78**, 620–625.

Lehmann, Erich L. and Casella, George (1998). *Theory of Point Estimation*, 2nd edn. Springer, New York.

Montgomery, Alan L., Zarnowitz, Victor, Tsay, Ruey S. and Tiao, George C. (1998). Forecasting the U.S. unemployment rate. *J. Amer. Stat. Assoc.*, **93**, 478–493.

Narasimham, Gorti V.L., Swamy P.A.V.B. and Reed, R.C. (1988). Productivity analysis of U.S. manufacturing using a stochastic-coefficients production function. *J. Busin. Econom. Stat.*, **6**, 339–350.

Oakes, D. (1985). Self-calibrating priors do not exist. *J. Amer. Stat. Assoc.*, **80**, 339.

Peddada, S.D. (1985). A short note on Pitman's measure of nearness. *Amer. Statistician*, **39**, 298–299.

Pratt, John W. and Schlaifer, Robert (1984). On the nature and discovery of structure. *J. Amer. Stat. Assoc.*, **79**, 9–21, 29–33.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd edn. Wiley, New York.

Schervish, M.J. (1985). Discussion of 'Calibration-based empirical probability' by A.P. Dawid. *Ann. Statist.*, **13**, 1274–1282.

Swamy, P.A.V.B. and Mehta, Jatinder S. (1975). Bayesian and non-Bayesian analysis of switching regressions and of random coefficient regression models. *J. Amer. Stat. Assoc.*, **70**, 593–602.

Swamy, P.A.V.B. and Schinasi, Garry J. (1989). Should fixed coefficients be re-estimated every period for extrapolation? *J. Forecasting*, **8**, 1–18.

Swamy, P.A.V.B., Mehta, Jatinder S. and Singamsetti, Rao N. (1996). Circumstances in which different criteria of estimation can be applied to estimate policy effects. *J. Statist. Planning Inference*, **50**, 121–153.

Swamy, P.A.V.B. and Tavlas, George S. (2001). Random coefficient models. In B.H. Baltagi (ed.), *Companion to Theoretical Econometrics*. Blackwell Publishers Ltd, Malden, Massachusetts, pp. 410–428.

Yokum, J. Thomas, Wildt, Albert R. and Swamy, P.A.V.B. (1998). Forecasting disjoint data structures using approximate constant and stochastic coefficient models. *J. Appl. Statist. Sci.*, **8**, 29–49.