Empirical best linear unbiased prediction in misspecified and improved panel data models with an application to gasoline demand

P.A.V.B. Swamy\textsuperscript{a,*}, Wisam Yaghi\textsuperscript{b}, Jatinder S. Mehta\textsuperscript{c}, I-Lok Chang\textsuperscript{d}

\textsuperscript{a}Statistical Methods Division, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics, Washington, DC 20212, USA
\textsuperscript{b}Department of Mathematics and Statistics, American University, Washington, DC 20016, USA
\textsuperscript{c}Department of Mathematics, Temple University, Philadelphia, PA 19122, USA
\textsuperscript{d}Department of Mathematics and Statistics, American University, Washington, DC 20016, USA

Available online 2 May 2006

Abstract

Misspecifications in econometric models can result in misestimated coefficients. An improved method for specifying econometric models is presented. The mean square error of an empirical best linear unbiased predictor of an individual drawing for the dependent variable of an improved model is derived. These ideas are illustrated using certain misspecified and improved models of the demand for gasoline in the US. It is shown that the forecasting gains from using the improved instead of the misspecified version of the gasoline demand model are very large. A description of a computational algorithm for combining iteratively re-scaled generalized least-squares estimation with out-of-sample multistep-ahead forecast generation is included.

Keywords: Correct interpretation of model coefficients; Forecasting performance

1. Introduction

Panel data sets can contain observations on many individuals, each observed at several points in time. Modeling in this setting calls for some general stochastic specifications since models must account for differences in behavior across individuals and changes in individual behavior over time as well as for specification bias. Specification bias occurs when relevant variables are omitted, when included explanatory variables are measured with error, or when incorrect functional forms are used. These biases cannot be analyzed with commonly used fixed- or random-effects specifications. This article presents some general panel-data models that can account for these biases as well as behavioral differences and changes.

Predictive testing—extrapolation to data outside the sample—is a good method for validating such models. It can be performed by computing the distributions of the absolute relative errors of out-of-sample forecasts (with a particular forecast horizon) generated from general stochastic specifications. This article also demonstrates how the mean square errors (MSEs) of such forecasts can be estimated. We apply this approach to a study of the demand for gasoline.
Section 2 specifies a general demand model for gasoline and uses the specified model to derive both a second-order approximation to the (average) MSE of an empirical best linear unbiased predictor (EBLUP) of a value of its dependent variable and an estimator of this approximate MSE to the desired order of approximation. Section 3 compares the forecasting performances of the misspecified and improved models using data on the US gasoline market. Section 4 concludes. Computational details are given in the Appendix.

2. Models of demand for gasoline

2.1. Models and assumptions

Economic theories specify that the quantity demanded of gasoline is a function of income and all prices. Accordingly, it may be specified that

\[ y_{it}^* = \beta_{0i} + \beta_{1i}x_{1it} + \beta_{2i}x_{2it}^* + \beta_{3i}x_{3it}^* + \beta_{4i}x_{4it}^* + \sum_{j=5}^{K_{it}} \beta_{j} x_{jit}^*, \]

(1)

where \( y_{it}^* \) is the log per capita gasoline consumption, computed as total gasoline consumption divided by a price index and population, \( x_{1it} \) the log price index for gasoline, \( x_{2it}^* \) the log per capita disposable income, \( x_{3it}^* \) the log price index for new cars, \( x_{4it}^* \) the log price index for used cars, \( K_{it} > 4 \), the \( x_{jit}^* \) with \( j > 4 \) are the remaining determinants of \( y_{it}^* \), \( i \) indexes individuals, \( t \) indexes time, and the symbols with an asterisk denote the “true” values. By definition, none of the variables of model (1) is mismeasured. To avoid excluding any determinant of \( y_{it}^* \) in Eq. (1) for any individual at any time, we assume that the number of the determinants of \( y_{it}^* \) may change over time for some or all individuals. Hence \( K_{it} \) is dependent on \( i \) and \( t \). In other words, there are no determinants of \( y_{it}^* \) excluded from (1).

In reality we do not have data for the \( x_{jit}^* \) with \( j > 4 \), and the data we do have for the other variables in (1) contain measurement errors. In what follows, symbols without an asterisk denote observable variables measured with error. Let \( y_{it}, x_{1it}, \ldots, x_{4it} \) denote the available observations on \( y_{it}^*, x_{1it}^*, \ldots, x_{4it}^* \), respectively. These observations are the sums of “true” values and measurement errors.

If we exclude from (1) the variables, say the \( x_{jit}^* \) with \( j > 4 \), then an explanation of the relationship between \( y_{it}^* \) and the \( x_{jit}^* \) with \( j < 5 \), can be found in the dependence of both \( y_{it}^* \) and the \( x_{jit}^* \) with \( j < 5 \) on the \( x_{jit}^* \) with \( j > 4 \), a phenomenon known as spurious correlation (see Lehmann and Casella, 1998, p. 107). Thus, by virtue of its including all the relevant explanatory variables, the formulation in (1) does not represent any spurious correlations.

If the demand model in (1) is rigorously deduced from maximization of the individual utility functions, then its functional form depends on the functional forms of the utility functions. Since the functional forms of the utility functions are unknown, imposing any undue restriction on the pattern of variation in the coefficients (e.g., constancy) may force (1) to have a wrong functional form. Thus, permitting all coefficients to vary freely produces an infinite class of functional forms that encompasses the correct one. The member of the class in (1) with the correct functional form depends on the functional forms of the utility functions. Since the functional forms of the utility functions are unknown, imposing any undue restriction on the pattern of variation in the coefficients (e.g., constancy) may force (1) to have a wrong functional form. Thus, permitting all coefficients to vary freely produces an infinite class of functional forms that encompasses the correct one. The member of the class in (1) with the correct functional form is given by the coefficients with the correct time profiles. These coefficients are called “the true coefficients” and are denoted by \( \beta_{jit}^* \), \( j = 0, 1, 2, \ldots, K_{it} \). Their existence is assumed here. Therefore, the “true” demand function for gasoline (or, simply, the “true” model) is

\[ y_{it}^* = \beta_{0i} + \sum_{j=1}^{4} \beta_{jit}^* x_{jit}^* + \sum_{j=5}^{K_{it}} \beta_{j} x_{jit}^*. \]

(2)

The unrealized values of \( y_{it}^* \) corresponding to the unrealized values of \( x_{jit}^* \), \( j = 1, 2, \ldots, K_{it} \), are called potential values if they are defined for every set of values of \( x_{jit}^* \), \( j = 1, 2, \ldots, K_{it} \), on every observation \( i \) and \( t \), whereas the realized values of \( y_{it}^* \) are those corresponding to the realized values of \( x_{jit}^* \), \( j = 1, 2, \ldots, K_{it} \). Only when these potential values exist is the “true” model a real-world relation and not a statistical association, as shown by Pratt and Schlaifer (1988, p. 28). If the potential values of \( y_{it}^* \) do not exist, the regression of \( y_{it} \) on \( x_{jit} \), \( j = 1, 2, 3, 4 \), fitted to observations is a pure statistical artifact.
To make the “true” model operational without misspecifying it, we proceed as follows: a freely-varying coefficient panel data (FVC-PD) model is

$$y_{it} = \gamma_{0it} + \sum_{j=1}^{4} \gamma_{jit} x_{jit}. \quad (3)$$

The most commonly used forms of this model are

$$y_{it} = \beta_{0} + \sum_{j=1}^{4} \beta_{j} x_{jit} + \nu_{0it}, \quad (4)$$

$$= \beta_{0} + \sum_{j=1}^{4} \beta_{j} x_{jit} + \mu_{t} + \lambda_{t} + \nu_{it}, \quad (5)$$

where \( \mu_{t} \) is an individual-specific, time-invariant fixed or random effect, \( \lambda_{t} \) a time-specific, individual-invariant fixed or random effect, and \( \nu_{it} \) the remainder disturbance. In the special case where accurate time-series data on the US gasoline market alone are available, model (4) is the same as the model considered by Greene (2003, p. 12). If our data on the \( x_{jit} \) with \( j < 5 \) were accurate, then Eq. (5) would be a two-way fixed or random effects model (see Greene, 2003, pp. 336–337). Models (4) and (5) treat certain features of the demand function in (1) as constant parameters. Years ago, Goldberger (1967) pointed out that this particular choice of constants might be questioned.

The variables, \( x_{jit} \), with \( j < 5 \) are called “the included explanatory variables” because they are included in the FVC-PD model. The variables, \( x_{jit}^{*} \), with \( j > 4 \) are called “excluded variables” because they are omitted from the FVC-PD model.

2.1.1. Correct interpretations of the coefficients of the FVC-PD model

Chang et al. (2000) prove that the sufficient conditions for the FVC-PD model to be an exact representation of the “true” model are (I) the intercept \( \gamma_{0it} \) is the sum of (i) \( x_{0it}^{*} \) in (2), (ii) the joint effect on \( y_{it}^{*} \) in (2) of the portions of excluded variables remaining after the effects of the “true” values, \( x_{jit}^{*} \), \( j = 1, 2, 3, 4 \), of the included explanatory variables have been removed, and (iii) the nonsampling and/or sampling errors in \( y_{it} \) in (3) and (II) for \( j = 1, 2, 3, 4 \), \( \gamma_{jit} \) is the sum of (i) \( x_{jit}^{*} \) in (2), (ii) a term capturing omitted-variables bias due to excluded variables, and (iii) a measurement-error bias due to mismeasuring \( x_{jit} \).

The omitted-variable bias components of the \( \gamma_{jit} \)'s are zero if the included explanatory variables are independent of “the” excluded variables themselves. But Pratt and Schlaifer (1988, p. 34) prove that this condition is meaningless unless the definite article is deleted and can then be satisfied only for certain “sufficient sets” of excluded variables. Since we do not know what these sufficient sets are, we assume that the omitted-variable biases are nonzero. The lack of perfect measurements on the included explanatory variables is the reason why the measurement-error bias components of the \( \gamma_{jit} \)'s are nonzero. The coefficients of the FVC-PD model will be called “the biased coefficients” because they contain omitted-variable and measurement-error biases. The coefficients of the “true” model will be called “the bias-free coefficients”, since they do not contain any biases.

The “true” model is not a real-world relation unless the potential values of its dependent variable exist, in which case models (4) and (5) are spurious if \( x_{jit}^{*} \), \( j = 1, \ldots, 4 \), are zero for all \( i \) and \( t \) and are otherwise not spurious, but could have an incorrect functional form. If an estimate of the coefficient of \( x_{jit} \) in (4) or (5) has a wrong sign, this is the result of improperly correcting for omitted-variables and measurement-error bias in the coefficient. Also, the significance (or insignificance) of an estimate of \( \gamma_{jit} \) in the FVC-PD model is not a good indicator of the significance (or insignificance) of the implied estimate of \( x_{jit}^{*} \) in the “true” model. For this reason, we suggest below a method for decomposing the coefficients of the FVC-PD model into their components.

2.1.2. Implications of the correct interpretations of the coefficients of the FVC-PD model

(i) Variation in \( \gamma_{jit} \) stems from variations in its components. The real-world sources of variations in these components are: (a) the nonlinearities of the “true” model causing variations in the \( x^{*} \)'s, (b) the nonlinearities of the relationships among the “true” values of excluded and included explanatory variables, (c) variation in the relative error of measurement...
in \(x_{jit}\), and (d) changes in the values of \(K_{it}\) as well as in the variables that constitute \(K_{it}\) over time and across individuals.

(ii) The measurement-error bias component of \(\gamma_{jit}\) is a function of both \(x_{jit}\) and the measurement error in \(x_{jit}\), implying that \(x_{jit}\) cannot be uncorrelated with its own coefficient, \(\gamma_{jit}\). (iii) For \(j = 0, 1, \ldots, 4\), the \(\gamma_{jit}\) cannot be uncorrelated with each other since all are dependent on the same coefficients, \(z_{jit}^*, j = 5, \ldots, K_{it}\), of excluded variables (see Swamy et al., 2003).

The assumptions under which models (4) and (5) are usually estimated are:

For \(j = 1, 2, 3, 4\) and all \(i\) and \(t\),

(A1) in the FVC-PD model, the distribution of \(\gamma_{jit}\) is degenerate at \(\beta_j\) and the distribution of \(\gamma_{0it}\) is the same as that of \(\beta_0 + \varepsilon_{0it}\);

(A2) the error term, \(\varepsilon_{0it}\), of model (4) follows a first-order autoregressive process, \(\varepsilon_{0it} = \phi_{0it}\varepsilon_{0it-1} + u_{0it}\), where \(|\phi_{0it}| < 1\), and the \(u_{0it}\) are independently distributed with mean zero and variance \(\sigma_{0it}^2\);

(A2') in model (5), the \(\mu_{it}\) are identically and independently distributed (i.i.d.) with mean zero and variance \(\sigma_{\mu_{it}}^2\); the \(\lambda_t\) are i.i.d. with mean zero and variance \(\sigma_{\lambda_t}^2\); the \(v_{it}\) are i.i.d. with mean zero and variance \(\sigma_{v_{it}}^2\), and \(\mu_{it}, \lambda_t, v_{it}\) are mutually independent;

(A3) the included explanatory variable \(x_{jit}\) is (mean) independent of \(\varepsilon_{0it}\);

(A3') some included explanatory variables and some excluded variables are exogenous or (mean) independent of \(\varepsilon_{0it}\).

None of the components of \(\gamma_{0it}\) in the FVC-PD model is equal to \(\sum_{jit=5}^{K_{it}} z_{jit}^* x_{jit}^*\) and hence the interpretation of the error term \(\varepsilon_{0it}\) of model (4) or the sum \(\mu_{it} + \lambda_t + v_{it}\) of model (5) as representing the net effect of the excluded variables on \(y_{jit}\) is incorrect. The correct interpretations of the coefficients of the FVC-PD model contradict Assumptions (A1) and (A3) (or (A3')) because the relative error of measurement in \(x_{jit}\) cannot be constant even if the other sources of variation in \(\gamma_{jit}\) are absent. With the measurement-error bias component of \(\gamma_{jit}\) varying, (A3) or (A3') cannot be true. This result shows that the label “the misspecified panel data (M-PD) models” is appropriate for models (4) and (5).

Rewriting (1) in terms of \(x_{jit}\), \(j = 1, 2, 3, 4\), and a function of \(x_{jit}\), \(j = 1, 2, 3, 4\), and \(x_{jit}\), \(j = 5, 6, \ldots, K_{it}\), leaves the coefficients of the FVC-PD model with the correct interpretations invariant (see Swamy et al., 1996). The coefficients of (1) and the M-PD models, however, do not possess this invariance property. Consequently, \(x_{jit}, j = 5, 6, \ldots, K_{it}\), and \(x_{jit}, j = 1, \ldots, 4\), in (1), and \(\beta_j, j = 0, 1, \ldots, 4\), and \(\varepsilon_{0it}\), in the M-PD models are not unique, as shown in Pratt and Schlaifer (1984, p. 13). By contrast, the “true” model (2) is unique if it is a real-world relation that remains invariant against mere changes in the language we use to describe it, as shown by Basman (1988, pp. 72–74). For example, both adding and subtracting a term on the right-hand side of a real-world relation changes only its representation but not the relation itself. The FVC-PD model, but not the M-PD models, shares this invariance property with the real-world relations.

Writing \(\varepsilon_{0it}\) as \(\mu_{it} + \lambda_t + v_{it}\), as in (5), does not limit the effects of the misspecifications of the M-PD model (4). This is because the decomposition \(\varepsilon_{0it}\) is \(\mu_{it} + \lambda_t + v_{it}\) is entirely arbitrary, since \(\varepsilon_{0it}\) itself is not unique.

The connections between the coefficients of the “true” model and those of the FVC-PD model pass undetected if the correct interpretations of the latter coefficients are unnoticed. If the correlation between \(x_{jit}\) and its coefficient \(\gamma_{jit}\) is ignored, if the distribution assumed for \(\gamma_{jit}\) is inconsistent with the “true” distributions of its components, or if the assumptions made about the initial values of \(\gamma_{jit}\), needed in Kalman-filter applications are inconsistent with the “true” distributions of its components, then the FVC-PD model can give poor predictions of \(y_{jit}\) and inconsistent estimators of \(z_{jit}^*, j = 0, 1, \ldots, 4\).

To obtain a forecast of \(y_{jit}\), we need a conditional mean of \(y_{jit}\) given some of its determinants because the average MSE of \(y_{jit}\) about a function of the given determinants is a minimum when the function is equal to the conditional mean, provided the average MSE is finite (see Rao, 1973, p. 264). The derivation of the conditional distribution of \(y_{jit}\) given the \(x_{jit}\) in the FVC-PD model is complicated by the fact that \(x_{jit}\) is correlated with its coefficient \(\gamma_{jit}\). To account for these correlations, we assume that for \(j = 0, 1, \ldots, 4\), \(i\) and \(t' = 1, \ldots, n\), and all \(t\) and \(t'\),

\[
\gamma_{jit} = \pi_{j0} + \sum_{d=1}^{p-1} \pi_{jd} z_{dit} + \mu_{ji} + \sum_{h=0}^{m-1} l_{jh} \varepsilon_{hit},
\]

where none of the \(z_{dit}\) is equal to 1 for all \(i\) and \(t\); the \(z_{dit}\) will be called “the coefficient drivers”; the \(l_{jh}\) are known positive constants; \(L = [l_{jh}]\) is a \(5 \times m\) matrix having \(l_{jh}\) as its \((j+1, h+1)\) element, \(m < , = , \) or \(> 5\),
(B2) the 5-vector $\mu_i = (\mu_{0i}, \mu_{1i}, \ldots, \mu_{4i})'$ is distributed with $E \left( \mu_i \mid z_{1it}, \ldots, z_{p-1it} \right) = 0$ and

$$E \left( \mu_i, \mu_i' \mid z_{1it}, \ldots, z_{p-1it} \right) = \begin{cases} \Delta & \text{if } i = i', \\ 0 & \text{if } i \neq i', \end{cases}$$

where $\Delta$ may not be diagonal,

(B3) the $m$-vector $e_{it} = (e_{0it}, e_{1it}, \ldots, e_{m-1it})'$ follows the stochastic equation

$$e_{it} = \phi_{ij} e_{i-1} + u_{it},$$

where $\phi_{ij}$ is an $m \times m$ (not necessarily diagonal) matrix whose eigenvalues are less than 1 in absolute value, the $m$-vector $u_{it} = (u_{0it}, u_{1it}, \ldots, u_{m-1it})'$ is distributed with $E \left( u_{it} \mid z_{1it}, \ldots, z_{p-1it} \right) = 0$ and

$$E \left( u_{it}, u_{i't} \mid z_{1it}, \ldots, z_{p-1it} \right) = \begin{cases} \sigma_{ii}^2 \Delta_{ii} & \text{if } i = i' \text{ and } t = t', \\ 0 & \text{if } i \neq i' \text{ and } t \neq t', \end{cases}$$

where $\Delta_{ii}$ may not be diagonal,

(B4) given $z_{dit}, d = 1, \ldots, p - 1,$ $\mu_i$ and $e_{it}$ are independent of each other and each varies independently across $i$, and

(B5) given $z_{dit}, d = 1, \ldots, p - 1,$ the $\mu_i$'s and $e_{it}$'s are independent of the $x_{jit}$.

We usually set $L = I$. The elements $\mu_{ji}$ are constant through time. They account for those between-individual variations in $\gamma_{jit}$, that are not explained by the $z_{dit}$ and $e_{hit}$ included in (6). The $e_{hit}$'s differ among individuals both at a given point in time and through time.

Assumptions (B1) and (B5) state that the correlation between $x_{jit}$ and $\gamma_{jit}$ is due to the mean function, $\pi_{j0} + \sum_{d=1}^{p-1} \pi_{jd} z_{dit}$, but once this function is subtracted from $\gamma_{jit}$, the remainder, $\mu_{ji} + \sum_{h=0}^{m-1} l_{jh} e_{hit}$, is independent of $x_{jit}$. Thus, “proper” coefficient drivers are those that are highly correlated with the included explanatory variables and uncorrelated with the $\mu_{ji}$ and $e_{hit}$. When proper coefficient drivers are found, (B1) and (B5) are consistent with the correct interpretations of the coefficients of the FVC-PD model. The FVC-PD model with coefficients satisfying Assumptions (B1)–(B5) is called “the stochastic coefficient panel data (SC-PD) model”.

### 2.1.3. A crucial distinction

The SC-PD model differs from the hierarchical Bayes model given in Lehmann and Casella (1998, pp. 253–262) in that the distribution of the coefficients of the former is part of the likelihood function, whereas the distribution of the coefficients of the latter is part of the prior distribution.

A vector formulation of the SC-PD model is

$$y_{it} = x_{it}' \gamma_{it},$$

where $x_{it} = (x_{0it}, \ldots, x_{4it})'$ is a 5-vector with $x_{0it} = 1$ for all $i$ and $t$ and having $x_{jit}$ as its $(j + 1)$th element, $\gamma_{it} = (\gamma_{0it}, \ldots, \gamma_{4it})'$ is a 5-vector having $\gamma_{jit}$ as its $(j + 1)$th element. A matrix formulation of (6) is

$$\gamma_{it} = \Pi z_{it} + \mu_i + L e_{it},$$

where $z_{it} = (z_{0it}, \ldots, z_{p-1it})'$ is a $p$-vector with $z_{0it} = 1$ for all $i$ and $t$ and having $z_{dit}$ as its $(d + 1)$th element, $\Pi = [\pi_{jd}]_{0 \leq j \leq 4, 0 \leq d \leq p - 1}$ is a $5 \times p$ matrix having $\pi_{jd}$ as its $(j + 1, d + 1)$ element, $\mu_i, L,$ and $e_{it}$ are as defined below (6). Substituting (11) into (10) gives

$$y_{it} = (z_{it}' \otimes x_{it}) \pi^{Long} + x_{it}' (\mu_i + L e_{it}), \quad t = 1, \ldots, T_i, \quad i = 1, \ldots, n,$$

where $\otimes$ denotes a Kronecker product, and $\pi^{Long}$ is a $5p$-vector, denoting a column stack of $\Pi$. We call (12) “the improved panel data (I-PD) model”.

The observations in (12) can be displayed in a matrix form as
\[ y_i = X_{zi} \pi_{Long} + X_i \mu_i + D_{xi} (I_{Ti} \otimes L) \hat{e}_i, \quad i = 1, \ldots, n, \] (13)
where \( y_i = (y_{i1}, \ldots, y_{iT_i})' \) is a \( T_i \)-vector, \( X_{zi} = (z_{i1} \otimes x_{i1}, \ldots, z_{iT_i} \otimes x_{iT_i})' \) is \( T_i \times 5p \), \( X_i = (x_{i1}, \ldots, x_{iT_i})' \) is \( T_i \times 5 \), \( D_{xi} = \text{diag}_{1 \leq t \leq T_i} (x_{it}')' \) is \( T_i \times 5T_i \), \( I_{ti} \) is an identity matrix of order \( T_i \), and \( \hat{e}_i = (\hat{e}_{i1}', \ldots, \hat{e}_{iT_i}')' \) is a \( T_i m \)-vector.

Under (B1)–(B5), the conditional expectation of \( y_i \) given \( X_{zi} \) is \( X_{zi} \pi_{Long} \) and the conditional covariance matrix of \( y_i \) given \( X_{zi} \) is \( X_{zi} \Delta X_i' + D_{xi} \left( I_{Ti} \otimes L \right) \sigma_i^2 \Sigma_{\hat{e}_i} \left( I_{Ti} \otimes L \right)' D_{xi}' \), denoted \( \Omega_{\hat{e}_i} \), where \( \sigma_i^2 \Sigma_{\hat{e}_i} \) is the covariance matrix of \( \hat{e}_i \). The matrix \( \Sigma_{\hat{e}_i} \) is \( T_i m \times T_i m \) having the \( m \times m \) matrix \( E \left( \hat{e}_{it} \hat{e}_{is}' \right) / \sigma_i^2 = \Gamma_{ii} / \sigma_i^2 = \phi_{ii} (\Gamma_{ii} / \sigma_i^2) \phi_{ii}' + \Delta_{ii} \) as its \( t \)-th diagonal sub-matrix, the \( m \times m \) matrix, \( E \left( \hat{e}_{it} \hat{e}_{is}' \right) / \sigma_i^2 = \phi_{ii}^{-1} (\Gamma_{ii} / \sigma_i^2) \phi_{ii}' \), as its \( (t, s) \)-sub-matrix with \( t > s \), and the \( m \times m \) matrix, \( E \left( \hat{e}_{it} \hat{e}_{is}' \right) / \sigma_i^2 = (\Gamma_{ii} / \sigma_i^2) (\phi_{ii}')_{s-t} \), as its \( (t, s) \)-sub-matrix with \( s > t \). Note that a stack of the \( n \) columns \( X_i \mu_i + D_{xi} \left( I_{Ti} \otimes L \right) \hat{e}_i \) has a block diagonal covariance structure.

### 2.1.4. Identification

In (13), \( \pi_{Long} \) and \( \mu_i \) are identifiable if \( X_{zi} \) and \( X_i \) have full column ranks, respectively, and \( \Omega_{\hat{e}_i} \) is nonsingular. The vector \( \hat{e}_i \) is not identifiable because \( D_{xi} \left( I_{Ti} \otimes L \right) \) does not have full column rank. We may be able to choose the \( z_{dit} \) in such a way that for \( j > 0 \), the sum of some of the first \( p \) terms on right-hand side of (6) is equal to the bias-free component of \( y_{jit} \). This result proves that only those bias-free coefficients of the “true” model that are also the components of the coefficients of the FVC-PD model are identifiable—subject to the restrictions imposed by (B1)—on the basis of the available data, whereas the bias-free coefficients on excluded variables are not identifiable.

Pratt and Schlaifer (1988, p. 49) produce a very convincing argument to show that a Bayesian will do much better to search like a non-Bayesian for concomitants that absorb omitted-variable and measurement-error biases. Using (B1), we do exactly as Pratt and Schlaifer suggest. The coefficient drivers in (6) are our concomitants. If they can decompose the coefficients of the FVC-PD model into the bias-free and omitted-variables and measurement-error bias components, then they should appear as the explanatory variables of the coefficients, as in (6). Note the distinction between the proper coefficient drivers and the econometrician’s instrumental variables.

### 2.2. A forecasting procedure

In this section, we consider the problem of estimating a linear combination \( \theta_i = w_i' \pi_{Long} + c_i' \mu_i + e_i' \hat{e}_i \) of the regression coefficients \( \pi_{Long} \) and the error components \( \mu_i \) and \( \hat{e}_i \), for the specified vectors, \( w_i, c_i, \) and \( e_i \), of constants by a linear estimator of the form \( \hat{\theta}_i = a_i' y_i + b_i \) for known \( a_i \) and \( b_i \). This linear estimator is unbiased for \( \theta_i \); that is, \( E \left( \hat{\theta}_i \right) = E \left( \theta_i \right) \), where \( E \) denotes the expectation with respect to (13), if and only if \( a_i' X_{zi} = w_i' \) and \( b_i = 0 \). The average MSE of \( \hat{\theta}_i \) about \( \theta_i \)
is given by $E\left(\hat{\theta}_i - \theta_i\right)^2$, where the expectation is taken with respect to variation in both $\hat{\theta}_i$ and $\theta_i$ implied by (13).

Minimizing $E\left(\hat{\theta}_i - \theta_i\right)^2$ subject to $a_i'X_{zi}^i = w_{i}^*$ gives $a_i'y_i = w_{i}^*\pi^{\text{Long}}(\omega) + c_i'\hat{\mu}_i(\omega) + e_i'\hat{\epsilon}_i(\omega)$, where

$$
\pi^{\text{Long}}(\omega) = \left(X_{zi}'\Omega_{ii}^{-1}X_{zi}\right)^{-1}X_{zi}'\Omega_{ii}^{-1}y_i,
$$

(14)

$$
\hat{\mu}_i(\omega) = \Delta X_i'\Omega_{ii}^{-1}\left\{y_i - X_{zi}\pi^{\text{Long}}(\omega)\right\},
$$

(15)

and

$$
\hat{\epsilon}_i(\omega) = \sigma_i^2\Sigma_{ii}\left(I_{T_i} \otimes L'\right)D_{xi}'\Omega_{ii}^{-1}\left\{y_i - X_{zi}\pi^{\text{Long}}(\omega)\right\},
$$

(16)

when $\omega$ is a known vector consisting of all the distinct nonzero elements of $\Lambda$, $\phi_{ii}$, $A_{ii}$, and $\sigma_i^2$. Predictor (16) exists, even though $D_{xi}'\left(I_{T_i} \otimes L\right)$ does not have full column rank.

Let $M_i = \left[I_{T_i} - X_{zi}'\Omega_{ii}^{-1}X_{zi}\right]^{-1}X_{zi}'\Omega_{ii}^{-1}$. Then it can be shown that if $\mu_i$ and $\epsilon_i$ are normal or the conditional expectations $E\left(\mu_i \mid M_iy_i\right)$ and $E\left(\epsilon_i \mid M_iy_i\right)$ are linear, then $\hat{\mu}_i(\omega) = E\left(\mu_i \mid M_iy_i\right)$ and $\hat{\epsilon}_i(\omega) = E\left(\epsilon_i \mid M_iy_i\right)$. The article by Swamy et al. (2003) exploits these properties to derive the estimators in (15) and (16). Another property of (15) and (16) shown in Swamy and Mehta (1975) is that they have minimum norm among all solutions to the equations, $M_iy_i = M_i\left(X_i, D_{xi}'\left(I_{T_i} \otimes L\right)\right)\left(\mu_i', \epsilon_i'\right)'$, when the norm of $\left(\mu_i', \epsilon_i'\right)'$ is defined as $\text{diag}\left[\left(\mu_i'\Lambda^{-1}\mu_i\right)^{1/2}, \left(\epsilon_i'\Sigma_{ii}^{-1}/\sigma_i^2\right)^{1/2}\right]$.

The results in (14)–(16) give the following $q$-step-ahead predictor of the value of $y_{iT_i+q}$, denoted by $y_{iT_i+q}$, from forecasting origin $T_i$ for the I-PD model:

$$
\hat{y}_{iT_i+q}(\omega) = \left(z_{iT_i+q}' \otimes x_{iT_i+q}'\right)\pi^{\text{Long}}(\omega) + x_{iT_i+q}'\left\{\hat{\mu}_i(\omega) + L\hat{\phi}_i^q\hat{\epsilon}_iT_i(\omega)\right\},
$$

(17)

where $q$ is a forecast horizon, $\hat{\epsilon}_iT_i = \sigma_i^2\Sigma_{ii}T_{iT_i}\left(I_{T_i} \otimes L'\right)D_{xi}'\Omega_{ii}^{-1}\left\{y_i - X_{zi}\pi^{\text{Long}}(\omega)\right\}$ with $\Sigma_{ii}T_{iT_i} = \left[\phi_i^T\left(\Gamma_{ii}/\sigma_i^2\right), \ldots, \Gamma_{ii}/\sigma_i^2\right]$ comprising the last $m$ elements of $\hat{\epsilon}_iT_i$ in (16). Predictor (17) is obtained by minimizing the average MSE of $a_i'y_i$ about $y_{iT_i+q} = \left(z_{iT_i+q}' \otimes x_{iT_i+q}'\right)\pi^{\text{Long}} + x_{iT_i+q}'\left(\mu_i + L\hat{\phi}_i^q\hat{\epsilon}_iT_i(\omega)\right)$ subject to $a_i'y_{iT_i+q} = \left(z_{iT_i+q}' \otimes x_{iT_i+q}'\right)\pi^{\text{Long}}$.

Hence it is the BLUP of $y_{iT_i+q}$ in the sense that it has minimum average MSE among all linear—in $y_i$—unbiased predictors of $y_{iT_i+q}$, when the I-PD model coincides with the “true” model.

In the case that $\omega$ is unknown, following an iteratively re-scaled generalized least squares (IRSGLS) method, Chang et al. (1992) and Swamy et al. (2003) estimate $\pi^{\text{Long}}$ and $\omega$ jointly from the observations in (13) subject to the restrictions that the eigenvalues of $\phi_{ii}$ are less than 1 in absolute value, and $\Lambda$ and $A_{ii}$ are nonnegative definite. Using the IRGSL estimator $\hat{\omega}$ of $\omega$ in place of $\omega$ used in (17) gives

$$
\hat{y}_{iT_i+q}(\hat{\omega}) = \left(z_{iT_i+q}' \otimes x_{iT_i+q}'\right)\pi^{\text{Long}}(\hat{\omega}) + x_{iT_i+q}'\left\{\hat{\mu}_i(\hat{\omega}) + L\hat{\phi}_i^q\hat{\epsilon}_iT_i(\hat{\omega})\right\},
$$

(18)

Following Rao (2003, p. 5), we call (18) “an empirical BLUP (EBLUP)” of $y_{iT_i+q}$.

An EBLUP of $y_{iT_i+q}$ constructed from the M-PD model in (4) under (A1)–(A3) is

$$
\tilde{y}_{iT_i+q}(\hat{\phi}_{0ii}) = x'_{iT_i+q}\tilde{\beta}\left(\hat{\phi}_{0ii}\right) + \hat{\phi}_{0ii}'\tilde{\epsilon}_{0iT_i}(\hat{\phi}_{0ii}),
$$

(19)

where $\tilde{\beta}(\hat{\phi}_{0ii})$ is an estimator of $\beta = (\beta_0, \beta_1, \ldots, \beta_q)'$ obtained by using an estimator, $\tilde{\phi}_{0ii}$, of $\phi_{0ii}$, in place of the known value of $\phi_{0ii}$ used in the GLS estimator of $\beta$ and $\tilde{\epsilon}_{0iT_i}(\hat{\phi}_{0ii})$ is the final element of $y_i - X_i\tilde{\beta}(\hat{\phi}_{0ii})$ (Greene, 2003, p. 279).

The I-PD model results in an improvement in the predictive ability over the M-PD model in (4) if

$$
\left|\frac{\hat{y}_{iT_i+q}(\hat{\omega}) - y_{iT_i+q}}{y_{iT_i+q}} - \left|\tilde{y}_{iT_i+q}(\hat{\phi}_{0ii}) - y_{iT_i+q}\right|/y_{iT_i+q}\right| < 0,
$$

(20)
with probability 1 as \( T_i \) varies for fixed \( q \) and each \( i \). The methods of estimation considered for the parameters of the M-PD model in (4) are maximum likelihood (ML) with the normality assumption about the \( u_{0it} \) in Assumption (A2), Yule–Walker (YW), iterative YW (IYW), unrestricted generalized least squares (UGLS), and restricted GLS (RGLS) with the restriction that \( \phi_{0ii} = 0 \).

It should be noted that incorrect results about the relative accuracies of (18) and (19) can be obtained if the MSEs, \( (1/F) \sum_{i=1}^{F-1} \left[ \hat{y}_{iT_i+q}(\hat{\theta}) - y_{iT_i+q} \right]^2 \) and \( (1/F) \sum_{i=1}^{F-1} \left[ \hat{y}_{iT_i+q}(\phi_{0ii}) - y_{iT_i+q} \right]^2 \) with \( F > 1 \), are used as measures of their inaccuracies, respectively. The article by Yokum et al. (1998) shows that, in the presence of structural changes represented by unexpected systematic model coefficient shifts superimposed on random variation, an I-PD model for \( y_{it} \) yields more accurate forecasts of \( y_{it} \) than a M-PD model for \( y_{it} \).

2.3. MSEs of BLUP and EBLUP

Using the elegant method in Rao (2003, pp. 98–99), we obtain the MSE of (17) about \( y_{iT_i+q} \):

\[
E\left[ \hat{y}_{iT_i+q}(\omega) - y_{iT_i+q} \right]^2 = E\left[ \left( \hat{y}_{iT_i+q}(\pi_{Long}, \omega) - y_{iT_i+q} + \hat{y}_{iT_i+q}(\omega) - \hat{y}_{iT_i+q}(\pi_{Long}, \omega) \right)^2 \right] = g_1(\omega) + g_2(\omega),
\]

where \( \hat{y}_{iT_i+q}(\pi_{Long}, \omega) \) is (17) with the “true” \( \pi_{Long} \) appearing in place of \( \hat{\pi}_{Long} \), \( g_1(\omega) = E\left[ \left( \hat{y}_{iT_i+q}(\omega) - y_{iT_i+q} \right)^2 \right] \) and \( g_2(\omega) \) measures the impact of sampling error in \( \hat{\pi}_{Long} \).

For (18), the prediction error can be written as

\[
\hat{y}_{iT_i+q}(\hat{\theta}) - y_{iT_i+q} = \left[ \hat{y}_{iT_i+q}(\omega) - y_{iT_i+q} \right] + \left[ \hat{y}_{iT_i+q}(\hat{\theta}) - \hat{y}_{iT_i+q}(\omega) \right].
\]

From this equation it follows that the MSE of \( \hat{y}_{iT_i+q}(\hat{\theta}) \) about \( y_{iT_i+q} \) is

\[
\text{MSE} \left[ \hat{y}_{iT_i+q}(\hat{\theta}) \right] = \text{MSE} \left[ \hat{y}_{iT_i+q}(\omega) \right] + E\left[ \left( \hat{y}_{iT_i+q}(\hat{\theta}) - \hat{y}_{iT_i+q}(\omega) \right)^2 \right].
\]

if \( E\left[ \left( \hat{y}_{iT_i+q}(\omega) - y_{iT_i+q} \right)^2 \right] = E\left[ \left( \hat{y}_{iT_i+q}(\hat{\theta}) - \hat{y}_{iT_i+q}(\omega) \right)^2 \right] = 0 \). Following Rao (2003, pp. 103 and 113–114), we can show that this condition is true if (i) \( \mu_i \) and \( \epsilon_i \) are normal, (ii) \( \hat{y}_{iT_i+q}(\hat{\theta}) - \hat{y}_{iT_i+q}(\omega) \) is a function of \( M_i y_i \), (iii) \( \left[ \frac{\hat{x}_{iT_i+q}}{\pi_{Long}(\omega) - \pi_{Long}} \right] \) is independent of \( M_i y_i \), (iv) \( x_{iT_i+q}^{\prime} \left[ \hat{\mu}_i(\omega) - \mu_i \right] = x_{iT_i+q}^{\prime} \left[ E \left( \hat{\mu}_i \right) \right] \), (v) \( x_{iT_i+q}^{\prime} \left[ \hat{\phi}_{ii}^p \epsilon_{iT_i+q} + \epsilon_{iT_i+q} \right] = x_{iT_i+q}^{\prime} \left[ \hat{\phi}_{ii}^p \right] \), (vi) \( x_{iT_i+q}^{\prime} \left[ \hat{\phi}_{ii}^p \right] = x_{iT_i+q}^{\prime} \left[ \hat{\phi}_{ii}^p \right] \)

We need to add the second term on the right-hand side of (22) to (21) to obtain the MSE of (18). The exact evaluation of this term is not possible and it is therefore necessary to obtain its approximation. Under the conditions of Theorem 1.1 in Lehmann and Casella (1998, p. 430), a Taylor approximation gives

\[
\hat{y}_{iT_i+q}(\hat{\theta}) - \hat{y}_{iT_i+q}(\omega) \approx \left\{ \frac{\partial \hat{y}_{iT_i+q}(\omega)}{\partial \omega} \right\}^{\prime}(\hat{\omega} - \omega),
\]

assuming that the terms involving higher powers of \( (\hat{\omega} - \omega) \) are of lower order relative to \( \left\{ \frac{\partial \hat{y}_{iT_i+q}(\omega)}{\partial \omega} \right\}^{\prime}(\hat{\omega} - \omega) \).

Under normality, \( \frac{\partial \hat{y}_{iT_i+q}(\omega)}{\partial \omega} \approx \frac{\partial \hat{y}_{iT_i+q}(\pi_{Long}(\omega))}{\partial \omega} = \left[ \frac{\partial x_{iT_i+q}^{\prime} \left( \pi_{Long}(\omega) - \pi_{Long}(\omega) \right)}{\partial \omega} \right] \Omega_{ii}^{-1}(\hat{\omega} - \omega) \) because the terms involving the derivatives of \( \hat{\pi}_{Long}(\omega) - \pi_{Long} \) with respect to \( \omega \) are of lower order. Thus,

\[
E\left[ \left\{ \frac{\partial \hat{y}_{iT_i+q}(\omega)}{\partial \omega} \right\}^{\prime}(\hat{\omega} - \omega) \right] = E\left[ \left\{ \frac{\partial x_{iT_i+q}^{\prime} \left( \pi_{Long}(\omega) - \pi_{Long}(\omega) \right)}{\partial \omega} \right\} \Omega_{ii}^{-1}(\hat{\omega} - \omega) \right]
\]
The third term on the right-hand side of this equation reduces to
approximately, but the bias of 
under (B1)–(B5) with
It has been noted by Rao (2003, p. 104) that
Again under the conditions of Theorem 1.1 in Lehmann and Casella (1998, p. 430), the bias of 
where \( V (\hat{\omega}) \) is the asymptotic covariance matrix of \( \hat{\omega} \). It now follows that
\[
E [\hat{y}_{iT_{i}+q} (\hat{\omega}) - \tilde{y}_{iT_{i}+q} (\hat{\omega})] \approx g_3 (\omega).
\]
Inserting (21) and (24) into (22) gives a second-order approximation to the MSE of (18) as
\[
\text{MSE} [\hat{y}_{iT_{i}+q} (\hat{\omega})] \approx g_1 (\omega) + g_2 (\omega) + g_3 (\omega).
\]
The terms \( g_2 (\omega) \) and \( g_3 (\omega) \) arise as the direct consequence of using the estimates of \( \pi^{\text{Long}} \) and \( \omega \) in (18), respectively. They are of lower order than the leading term, \( g_1 (\omega) \).

2.4. Estimation of the MSE of EBLUP

A naive approximation to the estimator of MSE \([\hat{y}_{iT_{i}+q} (\hat{\omega})]\) is obtained by substituting \( \hat{\omega} \) for \( \omega \) in (25):
\[
\text{MSE} [\hat{y}_{iT_{i}+q} (\hat{\omega})] \approx g_1 (\hat{\omega}) + g_2 (\hat{\omega}) + g_3 (\hat{\omega}).
\]
It has been noted by Rao (2003, p. 104) that \( E g_2 (\hat{\omega}) \approx g_2 (\omega) \) and \( E g_3 (\hat{\omega}) \approx g_3 (\omega) \) to the desired order of approximation, but the bias of \( g_1 (\hat{\omega}) \) is generally of the same order as \( g_2 (\omega) \) and \( g_3 (\omega) \).

Again under the conditions of Theorem 1.1 in Lehmann and Casella (1998, p. 430), the bias of \( g_1 (\hat{\omega}) \) can be evaluated using a Taylor expansion of \( g_1 (\hat{\omega}) \) around \( \omega \):
\[
g_1 (\hat{\omega}) = g_1 (\omega) + (\hat{\omega} - \omega)^T \nabla g_1 (\omega) + \frac{1}{2} (\hat{\omega} - \omega)^T \nabla^2 g_1 (\omega) (\hat{\omega} - \omega)
\]

\[
= g_1 (\omega) + \eta_1 + \eta_2,
\]
where \( \nabla g_1 (\omega) \) is the vector of first-order derivatives of \( g_1 (\omega) \) with respect to \( \omega \), \( \nabla^2 g_1 (\omega) \) is the matrix of second-order derivatives of \( g_1 (\omega) \) with respect to \( \omega \), \( \eta_1 = (\hat{\omega} - \omega)^T \nabla g_1 (\omega) \), and \( \eta_2 = \frac{1}{2} (\hat{\omega} - \omega)^T \nabla^2 g_1 (\omega) (\hat{\omega} - \omega) \). In IRSGLS estimation, the restriction that the estimates of \( \Lambda \) and \( \Lambda_{ij} \) are nonnegative definite is imposed. As a result, it is generally true that \( E \hat{\omega} \neq \omega \). Therefore, supposing that \( E (\eta_1) \) is of the same order as \( E (\eta_2) \) gives
\[
E g_1 (\hat{\omega}) \approx g_1 (\omega) + (E \hat{\omega} - \omega)^T \nabla g_1 (\omega) + \frac{1}{2} \text{tr} [\nabla^2 g_1 (\omega) V (\hat{\omega})].
\]
The third term on the right-hand side of this equation reduces to \( -g_3 (\omega) \) if \( \Omega_{ij} \) has a linear structure. Under this condition, an estimator of MSE \([\hat{y}_{iT_{i}+q} (\hat{\omega})]\) to the desired order of approximation is given by
\[
\text{MSE} [\hat{y}_{iT_{i}+q} (\hat{\omega})] \approx g_1 (\omega) - (\hat{E} \hat{\omega} - \hat{\omega})^T \nabla g_1 (\hat{\omega}) + g_2 (\hat{\omega}) + 2 g_3 (\hat{\omega}),
\]
where \( \hat{E} \hat{\omega} \) is an estimator of \( E \hat{\omega} \), since \( E \{ g_1 (\hat{\omega}) + g_3 (\hat{\omega}) \} - (E \hat{\omega} - \omega)^T \nabla g_1 (\omega) \approx g_1 (\omega) \), as shown by Rao (2003, p. 105). The term, \( (\hat{E} \hat{\omega} - \hat{\omega})^T \nabla g_1 (\hat{\omega}) \), in (27) can be ignored if \( (E \hat{\omega} - \omega)^T \nabla g_1 (\omega) \) is of lower order than \( E \eta_2 \).

3. Empirical results

In this section, (i) quarterly US data on \( y_{i_{11t}} \), \( x_{i_{11t}} \), \( j = 1, \ldots, 4 \), for 1953:1–2002:3 are used to estimate the FVC-PD model under (B1)–(B5) and the M-PD model in (4) under (A1)–(A3), (ii) the subscript \( i \) indexes only the US gasoline market, i.e., \( n = 1 \), (iii) \( L = I \), and (iv) \( \mu_t = 0 \), since only a single time series is considered. After some experimentation, we found that under (B1)–(B5) with \( x_{11t} = x_{11t-1} - x_{11t-2} \) and \( p = 2 \), the SC-PD model exhibited significantly improved...
Under (A1)–(A3), the ML estimates of the parameters of the M-PD model in (4) are wrong functional form. This we take to be the basic result of the specification error aspect of the M-PD model in (4). (A1) misspecifies the functional form of the “true” model and (A2) and (A3) are assumptions about a model with a functional form that is not support the assumption that the coefficients of the FVC-PD model are constant, particularly after 1974:1. Thus, in the SC-PD model, respectively. The estimates of the coefficients on log prices of new and used cars in the SC model lie within the ranges, of the corresponding coefficients in the M-PD model in (4) lie within these ranges, respectively. However, Fig. 1 shows that (i) the ML estimates of the coefficients on log gasoline price and log income in the M-PD model in (4) lie within the ranges, [−0.16, −0.07] and [0.48, 0.66], of the estimates of the corresponding coefficients in the SC-PD model, respectively. The estimates of the coefficients on log prices of new and used cars in the SC model with the ranges, [−0.06, 0.25] and [−0.08, 0.13], respectively, are positive for most of the quarters. The ML estimates of the corresponding coefficients in the M-PD model in (4) lie within these ranges, respectively. However, Fig. 1 does not support the assumption that the coefficients of the FVC-PD model are constant, particularly after 1974:1. Thus, (A1) misspecifies the functional form of the “true” model and (A2) and (A3) are assumptions about a model with a functional form. This we take to be the basic result of the specification error aspect of the M-PD model in (4).

Out-of-sample forecasts of $y_{it}$ are generated using predictors (18) and (19) for the I-PD and M-PD models in (12) and (4), respectively. Computational considerations are given in the Appendix. We use these forecasts to evaluate the difference between the absolute relative forecast errors for the I-PD and M-PD models defined in (20).

Table 1 shows that the percentage of the values of the difference in (20) that are less than or equal to 0.005 is 83.78, 79.28, 81.08, 77.48, and 72.07 for $q = 1, 2, 3, 4$, and 5, respectively. Although this percentage is high for all these values of $q$, it generally decreases as the forecast horizon $q$ increases. Further, it increases if the parameters of the M-PD model in (4) are estimated using the methods other than ML listed below (20).

The following conclusions emerge from our empirical results:

(i) The I-PD model based on (B1)–(B5) with $z_{it} = x_{1it} - x_{1i(t-2)}$, $L = I$, $\mu_i = 0$, and $p = 2$ results in a substantial improvement in the predictive ability over the M-PD model in (4) based on (A1)–(A3), as shown in line 2 of Table 1.
Table 1
Cumulative frequency distributions of the difference calculated as an absolute relative forecast error yielded by the I-PD model minus that yielded by the M-PD model in (4) for different forecast horizons

<table>
<thead>
<tr>
<th>Intervals of values</th>
<th>One-step ahead</th>
<th>Two-step ahead</th>
<th>Three-step ahead</th>
<th>Four-step ahead</th>
<th>Five-step ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 0</td>
<td>38.74</td>
<td>40.54</td>
<td>41.44</td>
<td>39.64</td>
<td>43.24</td>
</tr>
<tr>
<td>0.005</td>
<td>83.78</td>
<td>79.28</td>
<td>81.08</td>
<td>77.48</td>
<td>72.07</td>
</tr>
<tr>
<td>0.005–0.01</td>
<td>91.89</td>
<td>89.19</td>
<td>87.39</td>
<td>84.68</td>
<td>82.88</td>
</tr>
<tr>
<td>0.01–0.02</td>
<td>94.59</td>
<td>96.40</td>
<td>93.69</td>
<td>90.99</td>
<td>90.99</td>
</tr>
<tr>
<td>0.02–0.05</td>
<td>98.20</td>
<td>98.20</td>
<td>98.20</td>
<td>100.00</td>
<td>99.10</td>
</tr>
<tr>
<td>0.05–0.1</td>
<td>99.10</td>
<td>99.10</td>
<td>99.10</td>
<td>100.00</td>
<td>99.10</td>
</tr>
<tr>
<td>&gt; 0.1</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

(ii) The coefficient drivers other than $z_{1t}$ we considered assigned incorrect functional forms to the FVC-PD model and hence showed a deterioration of forecasting performance, especially for the multistep-ahead forecasts. The I-PD model is the best model only if it contains those coefficient drivers that assign the correct functional form to the FVC-PD model.

(iii) Misspecified models, whose underlying assumptions bear no resemblance to reality can predict well some of the time, but not all of the time, as shown in line 2 of Table 1.

(iv) The effects of the misspecifications of a model on its forecasting properties vary with the method used to estimate its parameters, as the effects of the different estimators of $\phi_{0li}$ on (20) show.

4. Conclusions

The biasing effects of measurement errors, omitted variables, and misspecifications of unknown functional forms are a pervasive problem in econometrics. We have illustrated an approach to account for such biases with the practical example of estimating the demand for gasoline in the US. For quarterly data on the US gasoline market, a stochastic coefficient model that adequately accounts for omitted-variable and measurement-error biases outperforms a misspecified model in terms of the forecasting accuracy for multistep-ahead forecasts. We have derived second-order approximations to the MSEs of out-of-sample forecasts generated from panel-data models with stochastic coefficients and their estimators to the desired order of approximation.

Acknowledgments

Any opinions expressed in this paper are those of the authors and do not constitute policy of their respective institutions. Thanks are due to J. N. K. Rao for helpful comments. Special thanks go to David Belsley and the two referees whose editorial changes have vastly improved the paper.

Appendix. Computational considerations

An IRSGLS method is used to estimate the parameters of the I-PD model and all the methods described below (20) are used to estimate the parameters of the M-PD model in (4). Out-of-sample forecasts of $y_{it}$ are generated from these two estimated models using a rolling forecast method. First, the model parameters are estimated using all observations through a given forecasting origin, $T_i$; next, the one- to five-quarter-ahead forecasts are generated for this origin using (18) and (19) with $q = 1, 2, \ldots, 5$, respectively. This procedure is then repeated for all forecasting origins beginning with 1973:4 and ending with 2001:2. The number of one- to five-quarter-ahead forecasts from the I-PD model or the M-PD model in (4) given by this method is 111.

The first step in computing the IRSGLS estimates of the parameters of the I-PD model for a given value of $T_i$ in the rolling forecast method is carried out using software developed by Chang et al. (2000). The appendix to Swamy
et al. (2005) describes some of the basic elements of computing the IRSGLS estimates and a number of tools that are used in this computation. Once the IRSGLS estimates are obtained, the evaluation of (18) using these estimates is straightforward.

References