# Latent Variable Autoregression with Exogenous Inputs

Daniil Bargman\*1

<sup>1</sup>Institute of Finance and Technology, University College London

June 26, 2025

#### **Abstract**

This paper introduces a new least squares regression methodology called (C)LARX: a (constrained) latent variable autoregressive model with exogenous inputs. Two additional contributions are made as a side effect: First, a new matrix operator is introduced for matrices and vectors with blocks along one dimension; Second, a new latent variable regression (LVR) framework is proposed for economics and finance. The empirical section examines how well the stock market predicts real economic activity in the United States. (C)LARX models outperform the baseline OLS specification in out-of-sample forecasts and offer novel analytical insights about the underlying functional relationship.

### 1 Introduction

Latent variables are unobserved processes which can be approximated from the observed data using statistical methods. Two popular latent variable models in modern econometrics are instrumental variable regression (e.g., Stock and Trebbi (2003)) and hidden Markov models (Baum and Petrie (1966)). In the former, the causal relationship between the explanatory and the dependent is obscured by an unobserved common factor. In the latter, the variables and/or their relationships are influenced by unobserved changes in regime such as the stage of the busniess cycle or a bull/bear market.

A somewhat different latent variable modelling paradigm has developed over the past few decades outside of economics and finance. Latent variable regression (LVR) models based on Canonical Correlation Analysis (CCA) (Hotelling (1936)) and Partial Least Squares (PLS) (Wold (1982, 1975)) take the concept of Principal Component Analysis (PCA) (Pearson (1901); Hotelling (1933)) into an inferential setting. Empirical measurements are viewed as partial or imperfect representations of the processes under observation. Linear combinations are constructed from those measurements in an attempt to better approximate the unobserved "true" processes and the relationships between them.

Over the past 90 years, the direction of CCA- and PLS-style LVR research (from now on simply LVR research) has been largely tangential to finance and economics, with the few identifiable counter-examples limited to advanced arbitrage pricing theory (e.g., see Bai and Ng (2006); Ahn et al. (2012)) and joint production functions (Vinod (1976, 1968)). Meanwhile, a large body of LVR literature has developed in industrial chemistry (e.g. Burnham et al. (1996, 1999)), machine learning (e.g., Wang et al. (2020); Dai et al. (2020); Van Vaerenbergh et al. (2018); Chi et al. (2013)), medicine and other quantitative fields (e.g., see Uurtio et al. (2017)). Traditionl use cases of LVR models now include dimensionality reduction, identification of the directions of correlation in multivariate data streams (e.g., Burnham et al. (1996); Dong and Qin (2018); Qin (2021)), and multi-label classification of images, videos, audio, and hand-written text (e.g., Wang et al. (2020); Dai et al. (2020); Van Vaerenbergh et al. (2018); Chi et al. (2013)).

This paper sets out to take LVR research in a direction more compatible with finance and economics applications. A new LVR methodology called (C)LARX – (constrained) latent variable autoregression

<sup>\*</sup>daniil.bargman.22@ucl.ac.uk

with exogenous inputs – is developed based on the same mathematical concepts that underpin LVR research in other disciplines. The (C)LARX methodology is designed as a superset of the ubiquitous ARX model in which any or all variables are allowed to be latent, i.e., to be represented by linear combinations over the observed data. All regression models in the (C)LARX family present with fixed point solutions interpretable through the lens of least squares regression and portfolio optimisation.

As this paper proceeds to show, (C)LARX and similar LVR models can address at least two practical use cases in economics and finance not fully covered by other regression techniques. First of all, these models allow the researcher to improve the accuracy of statistical measurement for the input variables based on the relationship between them. Second, the researcher can construct linear combinations of variables optimised for a specific functional dependency, e.g., with one linear combination being a leading indicator for the other. Both use cases have far-reaching implications for macroeconomics, investment management and beyond.

The rest of the paper is organised as follows. Section 2 defines a new matrix operator which is subsequently used to derive the fixed point solution for the (C)LARX family of models. Section 3 examines the underlying principles of LVR modelling in other disciplines and lays the foundation for an LVR framework in economics and finance. Section 5 derives the (C)LARX methodology. A brief overview of a few special cases of (C)LARX follows in Section 6. Section 7 presents a stylised empirical application of the (C)LARX model by examining the predictive power of the US stock market with respect to US economic activity. Section 8 concludes and discusses possible avenues for future research.

# 2 Blockwise Direct Sum Operator

The derivation of the (C)LARX fixed point in Section 5 involves blockwise operations between matrices and vectors with blocks along one dimension, including a blockwise Kronecker product  $\odot$  as defined in Khatri and Rao (1968). This paper also introduces a new matrix operation called a *blockwise direct sum*, denoted by the superscript  $\mathbf{A}^{\oplus}$  for an arbitrary matrix or vector  $\mathbf{A}$ . This section briefly introduces the blockwise direct sum operator (alternatively, a blockwise diagnoalisation operator) and its relevant properties.

Let  $\langle \mathbf{A}_i | 1 \leq i \leq k \rangle$  be a sequence of k real matrices with arbitrary dimensions. If all  $\mathbf{A}_i$  have the same number of columns, they can be concatenated vertically into a matrix with row blocks. If all  $\mathbf{A}_i$  have the same number of rows, they can be concatenated horizontally into a matrix with column blocks. Formally:

$$\langle \mathbf{A}_i \rangle \equiv \langle \mathbf{A}_i | 1 \leq i \leq k \rangle$$
 a sequence of matrices of arbitrary dimensions 
$$[\langle \mathbf{A}_i \rangle]_v = \mathbf{A}_v = \mathbf{A}$$
 a vertical matrix concatenation of  $\langle \mathbf{A}_i \rangle$  
$$[\langle \mathbf{A}_i \rangle]_h = \mathbf{A}_h$$
 a horizontal matrix concatenation of  $\langle \mathbf{A}_i \rangle$  
$$\langle \mathbf{M} \rangle$$
 the sequence of blocks comprising block matrix  $\mathbf{M}$ 

The blockwise direct sum operator  $\mathbf{A}^{\oplus}$  is then defined as:

$$\mathbf{A}^{\oplus} = \langle \mathbf{A} 
angle^{\oplus} = \langle \mathbf{A}_i 
angle^{\oplus} = \mathbf{A}_1 \oplus \mathbf{A}_2 \oplus \mathbf{A}_3 \oplus \cdots \oplus \mathbf{A}_k = egin{pmatrix} \mathbf{A}_1, & \mathbf{0}, & \cdots & \mathbf{0} \ \mathbf{0}, & \mathbf{A}_2, & \cdots & \mathbf{0} \ dots & dots & \ddots & dots \ \mathbf{0}, & \mathbf{0}, & \cdots & \mathbf{A}_k \end{pmatrix}$$

Compatibility of dimensions is not a constraint for a direct sum between matrices, so a blockwise direct sum is defined for any sequence of matrices  $\langle \mathbf{A}_i \rangle$ . The result is always a matrix whose number of rows (columns) is the sumtotal number of rows (columns) across all comprising  $\mathbf{A}_i$ .

The matrix  $\mathbf{A}^{\oplus}$  can itself be mapped to either a sequence of k row blocks or a sequence of k column blocks without slicing through the original matrices in the sequence. For most use cases the block structure of  $\mathbf{A}^{\oplus}$  will not be relevant. For the remaining scenarios let us apply the same shorthand notation as above:

$$\mathbf{A}^{\oplus} \equiv \mathbf{A}_v^{\oplus} = egin{pmatrix} \mathbf{A}_1, & \mathbf{0}, & \cdots & \mathbf{0} \ \hline \mathbf{0}, & \mathbf{A}_2, & \cdots & \mathbf{0} \ \hline dots & dots & \ddots & dots \ \hline \mathbf{0}, & \mathbf{0}, & \cdots & \mathbf{A}_k \end{bmatrix}, \quad \mathbf{A}_h^{\oplus} = egin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \ dots & dots & \ddots & dots \ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_k \end{pmatrix}$$

One of the main use cases for the blockwise direct sum operator lies in defining blockwise inner products and blockwise quadratic forms between matrices and vectors. For example, let  $\mathbf{B}$  be a matrix with the same number of rows and the same block structure as  $\mathbf{A}$ . The product  $(\mathbf{A}^{\oplus})'\mathbf{B}$  then yields:

$$\left(\mathbf{A}^{\oplus}
ight)'\mathbf{B} = egin{pmatrix} \mathbf{A}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_k' \end{pmatrix} egin{pmatrix} \frac{\mathbf{B}_1}{\mathbf{B}_2} \\ \hline \vdots \\ \hline \mathbf{B}_k \end{pmatrix} = egin{pmatrix} \frac{\mathbf{A}_1'\mathbf{B}_1}{\mathbf{A}_2'\mathbf{B}_2} \\ \hline \vdots \\ \hline \mathbf{A}_k'\mathbf{B}_k \end{pmatrix}$$

As showcased in subsequent chapters, this operation is useful for solving Lagrangian optimisation problems with respect to a vector of coefficients in the presence of peacemeal constraints, e.g., when various slices of the coefficient vector must each have unit length or zero-sum elements.

Several properties of the blockwise direct sum operator are relevant for this paper. First of all, we note that the transpose of  $A^{\oplus}$  is the same as the blockwise direct sum of A':

**Proposition 2.1.** The function composition of the blockwise direct sum operator and the transpose operator is commutative. In other words,  $(\mathbf{A}^{\oplus})' = (\mathbf{A}')^{\oplus}$ .

$$\left(\mathbf{A}^{\oplus}
ight)' = egin{pmatrix} \mathbf{A}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2' & \cdots & \mathbf{0} \\ dots & dots & \ddots & dots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_k' \end{pmatrix} = \mathbf{A}_1' \oplus \mathbf{A}_2' \oplus \mathbf{A}_3' \oplus \cdots \oplus \mathbf{A}_k' = \left(\mathbf{A}'\right)^{\oplus}$$

Second, the blockwise direct sum of a vector can be written as a blockwise Kronecker product:

**Proposition 2.2.** Let **a** comprise k blocks given by the sequence of vectors  $\langle \mathbf{a}_i | 1 \leq i \leq k \rangle$ . The matrix  $\mathbf{a}^{\oplus}$  can be expressed as a blockwise Kronecker product between **a** and an identity matrix  $I_k$  with vector blocks along the same dimension as **a**.

*Proof.* For a column vector we have:

$$\mathbf{a}^{\oplus} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{a}_k \end{pmatrix} = \begin{pmatrix} \mathbf{\underline{a}}_1 \otimes \begin{bmatrix} 1, & 0, & \cdots & 0 \end{bmatrix} \\ \hline \mathbf{\underline{a}}_2 \otimes \begin{bmatrix} 0, & 1, & \cdots & 0 \end{bmatrix} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{\underline{a}}_k \otimes \begin{bmatrix} 0, & 0, & \cdots & 1 \end{bmatrix} \end{pmatrix} = \begin{pmatrix} \mathbf{\underline{a}}_1 \\ \hline \mathbf{\underline{a}}_2 \\ \hline \vdots \\ \hline \mathbf{\underline{a}}_k \end{pmatrix} \odot \begin{pmatrix} \frac{1, & 0, & \cdots & 0}{0, & 1, & \cdots & 0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0, & 0, & \cdots & 1 \end{pmatrix} = \mathbf{a} \odot I_k$$

The proof for a row vector follows by symmetry.

Third, for two column vectors  $\mathbf{a}$  and  $\mathbf{b}$  with the same length and row block structure, the operation  $(\mathbf{a}^{\oplus})'\mathbf{b}$  is symmetric:

**Proposition 2.3.** For column vectors  $\mathbf{a}$  and  $\mathbf{b}$  with identically sized row blocks  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_k$  and  $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \dots, \mathbf{b}_k$ , respectively,  $(\mathbf{a}^{\oplus})' \mathbf{b} = (\mathbf{b}^{\oplus})' \mathbf{a}$ .

Proof.

$$\left(\mathbf{a}^{\oplus}
ight)' oldsymbol{b} = egin{pmatrix} \mathbf{a}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_2' & \cdots & \mathbf{0} \\ dots & dots & \ddots & dots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{a}_k' \end{pmatrix} egin{pmatrix} \frac{oldsymbol{b}_1}{oldsymbol{b}_2} \\ \hline dots \\ oldsymbol{b}_k \end{pmatrix} = egin{pmatrix} \frac{\mathbf{a}_1' oldsymbol{b}_1}{\mathbf{a}_2' oldsymbol{b}_2} \\ \hline dots \\ \hline oldsymbol{a}_k' oldsymbol{b}_k \end{pmatrix} = egin{pmatrix} \frac{oldsymbol{b}_1' oldsymbol{a}_1}{oldsymbol{b}_2' oldsymbol{a}_2} \\ \hline dots \\ \hline oldsymbol{b}_k' oldsymbol{a}_k \end{pmatrix} = egin{pmatrix} \frac{oldsymbol{b}_1' oldsymbol{a}_1}{oldsymbol{b}_2' oldsymbol{a}_2} \\ \hline dots \\ \hline oldsymbol{b}_k' oldsymbol{a}_k \end{pmatrix} = egin{pmatrix} oldsymbol{b}_1' oldsymbol{a}_1 \\ \hline dots \\ \hline oldsymbol{b}_k' oldsymbol{a}_k \end{pmatrix}$$

Lastly, the blockwise direct sum operator is commutative with respect to a certain class of operations over matrix sequences. Specifically, for a sequence of matrices  $\langle \mathbf{A}_i | 1 \leq i \leq k \rangle$  and an operation f over matrix sequences of length k, it can be shown that  $f(\langle \mathbf{A} \rangle)^{\oplus} = [f(\langle \mathbf{A}^{\oplus} \rangle)]$  if f satisfies certain conditions. This, in turn, can be used to prove that for two vectors  $\mathbf{a}$  and  $\mathbf{b}$  with k row blocks each, the blockwise Kronecker product  $\mathbf{a} \odot \mathbf{b}$  can be factorised in the same way as the traditional Kronecker product, namely:

$$\mathbf{a} \odot \mathbf{b} = (\mathbf{a} \odot I_{\mathbf{b}}) \mathbf{b} = (I_{\mathbf{a}} \odot \mathbf{b}) \mathbf{a}$$

where  $I_{\mathbf{a}}$  and  $I_{\mathbf{b}}$  are identity matrices with the same row block structure as  $\mathbf{a}$  and  $\mathbf{b}$ , respectively. The corresponding derivations are deferred to Appendices A and B.

# 3 A Latent Variable Regression Framework for Economics and Finance

Latent variable regression (LVR) models use statistical data to achieve two simultaneous objectives: to estimate a functional relationship, and to approximate one or more latent variables (LVs). For the second objective it is assumed that each LV can be approximated as a weighted sum over some sequence of observed data series. Formally:

**Definition 3.1.** A latent variable  $\tilde{y}$  is an unobserved process which can be approximated as a linear combination (weighted sum) over a vector of observed random variables  $Y = \begin{bmatrix} y_1, y_2, y_3, \dots, y_n \end{bmatrix}$  with two or more weights different from zero.

**Definition 3.2.** A latent variable model is a statistical technique aimed at estimating a latent variable weight vector  $\mathbf{w} = \begin{bmatrix} w_1, & w_2, & w_3, & \dots, & w_n \end{bmatrix}'$  such that  $\tilde{y} = Y\mathbf{w}$ , where  $\mathbf{w}$  can't be a multiple of a standard basis vector.

**Definition 3.3.** A latent variable regression (LVR) model is a regression model in which at least one variable is latent.

A traditional univariate regression model with dependent variable y, explanatory variable x and a vector of regression parameters  $\gamma$  can be defined as:

$$y = F_{\gamma}(x) + \epsilon \tag{1}$$

Here,  $\epsilon$  is a mean-zero error term and  $F_{\gamma}: \mathbb{R} \longrightarrow \mathbb{R}$  is generally assumed to be at least once differentiable with respect to  $\gamma$ . For illustrative purposes, let us further assume that (1) represents an equation for estimating the price elasticity of demand. This means that y might represent changes in the sale volume of a product, x the changes in its price, and  $F_{\gamma}$  would likely be some type of monotonically decreasing function.

A popular way to solve (1) is by finding a vector  $\hat{\gamma}$  which minimizes the variance of  $\epsilon$  for a given dataset of prices and sales volumes – a methodology known as least squares regression. In a slight

abuse of notation, let  $F(\mathbf{a})$  represent a row-wise operation over the elements in  $\mathbf{a}$  when  $\mathbf{a}$  is a sample vector or matrix. In other words, given a sample vector of observations  $\mathbf{a} = [a_1, a_2, a_3, \dots, a_s]'$  where s denotes the sample size, let  $F(\mathbf{a})$  represent  $[F(a_1), F(a_2), F(a_3), \dots, F(a_s)]'$ . Then, if we denote the sample observation vectors for y and x by y and x, respectively, we can write our least squares optimisation problem as:

$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmin}} \|\mathbf{y} - F_{\gamma}(\mathbf{x})\|_{2}^{2}$$
(2)

Having solved for  $\hat{\gamma}$ , we can look at how well  $F_{\hat{\gamma}}(\mathbf{x})$  predicts  $\mathbf{y}$ , preferably using a different sample of observations than the one to train the model. If the predictions are poor, we conclude that  $F_{\hat{\gamma}}$  is a poor approximation of the true price elasticity of demand.

With this "traditional" approach, no distinction is made between the quality of the model and the quality of the input data. However, in reality the sample vectors  $\mathbf{y}$  and  $\mathbf{x}$  may simply be inaccurate approximations of the product's true sale quantity and price: the same product may be offered at different prices in different locations, and the turnover data may come from different companies with different reporting rules. If the quality of the sample vectors  $\mathbf{y}$  and  $\mathbf{x}$  is in question, both y and x can be more accurately thought of as latent variables. We can represent this by rewriting equation (1) as an LVR model of the form:

$$\tilde{y} = F_{\gamma}(\tilde{x}) + \epsilon \tag{3}$$

According to Definition 3.2, if more than one source of information is available about the product's price ( $[x_1, x_2, x_3, \ldots, x_m] = X$ ) and its sales quantity ( $[y_1, y_2, y_3, \ldots, y_n] = Y$ ), equation (3) can be approximated as:

$$Y\mathbf{w} = F_{\gamma}(X\boldsymbol{\omega}) + \epsilon \tag{4}$$

If the solution to the traditional regression model (1) is given by the vector  $\hat{\gamma}$  which satisfies (2), then the solution to to the LVR model (4) would be given by the vectors  $\hat{\gamma}, \hat{\mathbf{w}}, \hat{\boldsymbol{\omega}}$  which satisfy:

$$\hat{\boldsymbol{\gamma}}, \hat{\mathbf{w}}, \hat{\boldsymbol{\omega}} = \underset{\boldsymbol{\gamma}, \mathbf{w}, \boldsymbol{\omega}}{\operatorname{argmin}} \|\mathbf{Y}\mathbf{w} - F_{\boldsymbol{\gamma}}(\mathbf{X}\boldsymbol{\omega})\|_{2}^{2}$$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_{1}, & \mathbf{y}_{2}, & \mathbf{y}_{3}, & \dots, & \mathbf{y}_{n} \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{1}, & \mathbf{x}_{2}, & \mathbf{x}_{3}, & \dots, & \mathbf{x}_{m} \end{bmatrix}$$
(5)

where  $\mathbf{y}_i$  and  $\mathbf{x}_j$  are column vectors with sample measurements for the respective  $y_i$  and  $\mathbf{x}_j$ .

There is, of course, a big conceptual difference between equations (1) and (3). In traditional regression models, y and x are observed a priori and their values are fixed when the relationship is being estimated. In LVR models, on the other hand, the very nature of  $\tilde{y}$  and  $\tilde{x}$  is determined by the functional relationship between them. The choice of F directly affects the computed values of  $\mathbf{w}$  and  $\boldsymbol{\omega}$ , so our estimates for both the sale quantity and the price will change depending on whether we choose to model the elasticity of demand as linear, quadratic or some other function. From a purely mathematical viewpoint, however, (5) is simply a generalization of (2):

**Proposition 3.1.** Given a non-empty vector of observed variables

$$\underset{1\times l}{A} = \begin{bmatrix} a_1, & a_2, & \dots, & a_l \end{bmatrix}$$

and a vector of linear combination weights

$$\mathbf{b} = \begin{bmatrix} b_1, & b_2, & \dots, & b_l \end{bmatrix}',$$

variable  $\tilde{a} = A\mathbf{b}$  can only be latent if l > 1.

*Proof.* Because A is non-empty and l is the number of columns in A, l must be a positive integer. The only case not covered by l > 1 is l = 1. If l = 1, both A and **b** can only have one (non-zero) element, which means that  $\tilde{a}$  is no longer a latent variable according to Definition 3.1.

Corollary 3.1.1. A latent variable regression of  $\tilde{y}$  on one or more  $\tilde{x}_j$  reduces to a traditional regression model when the underlying observed variable vectors Y and  $X_j$  each consist of a single variable.

*Proof.* If Y and  $X_j$  each have one element, it follows from Proposition 3.1 that none of  $\tilde{y}$ ,  $\tilde{x}_j$  are latent. Hence, a regression of  $\tilde{y}$  on  $\tilde{x}_j$  is not a latent variable regression according to Definition 3.3.

Because LVR models are a superset of traditional regression models, a common methodological framework can be defined for both. Assume we have an observed random variable space made up of:

- $\circ n \geq 1$  proxy measurements for the dependent variable, represented by a  $1 \times n$  row vector Y
- $M \ge 1$  proxy measurements for  $K \le M$  explanatory variables, represented by  $1 \times m_j$  row vectors  $X_j$  with j = 1, 2, ..., K and  $\sum_{j=1}^K m_j = M$

Over that space, define (latent or non-latent) variables  $\tilde{y} = Y_{\substack{\mathbf{w} \\ n \times 1}}$  and  $\tilde{x}_j = X_j \underbrace{\boldsymbol{\omega}_j}_{m_j \times 1}$ ,  $j = 1, 2, \dots, K$  governed by some functional relationship  $F_{\boldsymbol{\gamma}} : \mathbb{R}^K \longrightarrow \mathbb{R}$ , where  $\underset{p \times 1}{\boldsymbol{\gamma}}$  represents an unknown vector of regression parameters, such that:

$$\tilde{y} = F_{\gamma} (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \dots, \tilde{x}_K) + \epsilon 
\tilde{y} \equiv Y \mathbf{w} 
\tilde{x}_j \equiv X_j \boldsymbol{\omega}_j \text{ for } j = 1, 2, 3, \dots, K$$
(6)

The vectors  $\mathbf{w}$ ,  $\boldsymbol{\omega}_j$  and  $\boldsymbol{\gamma}$  are all initially unknown and need to be estimated empirically. This paper will continue to focus on the least squares approach which translates into the following objective function:

$$\|\tilde{y} - F_{\gamma}(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \dots, \tilde{x}_K)\|_2^2$$

$$(7)$$

With problems of this kind it is often necessary to impose constraints on  $\mathbf{w}$ ,  $\boldsymbol{\omega}$  and/or  $\gamma$ , e.g., to avoid trivial solutions such as  $\mathbf{w} = \mathbf{0}$ . For example, PLS-style LVR models (Wold (1975, 1982)) impose a unit length constraint on  $\mathbf{w}$  and  $\boldsymbol{\omega}$ , while CCA-style LVR models (Hotelling (1936)) impose a unit variance constraint on  $\mathbf{Y}\mathbf{w}$  and  $\mathbf{X}\boldsymbol{\omega}$ . We can generalize this to an arbitrary set of constraint functions  $\{g_i \colon \mathbb{R}^{M+n+p} \longrightarrow \mathbb{R}, i \in \mathbb{N}\}$ , where p is the length of  $\gamma$ .

The goal, then, is to estimate the vectors  $\hat{\gamma}$ ,  $\hat{\mathbf{w}}$  and  $\hat{\boldsymbol{\omega}}_j$  which minimize the variance of  $\epsilon$  subject to  $\{g_i \colon i \in \mathbb{N}\}$ , using a sample of s empirical observations given by the matrix  $\begin{bmatrix} \mathbf{Y} \\ s \times n \end{bmatrix}$ ,  $\begin{bmatrix} \mathbf{X}_1 \\ s \times m_1 \end{bmatrix}$ ,  $\begin{bmatrix} \mathbf{X}_2 \\ s \times m_2 \end{bmatrix}$ ,  $\begin{bmatrix} \mathbf{X}_1 \\ s \times m_2 \end{bmatrix}$ ,  $\begin{bmatrix} \mathbf{X}_2 \\ s \times m_2 \end{bmatrix}$ .

$$\min_{\boldsymbol{\gamma}, \mathbf{w}, \{\boldsymbol{\omega}_j\}} \|\mathbf{Y}\mathbf{w} - F_{\boldsymbol{\gamma}} (\mathbf{X}_1 \boldsymbol{\omega}_1, \mathbf{X}_2 \boldsymbol{\omega}_2, \mathbf{X}_3 \boldsymbol{\omega}_3, \dots, \mathbf{X}_K \boldsymbol{\omega}_K) \|_2^2, \quad j = 1, 2, \dots, K$$
s.t.  $g_i (\boldsymbol{\gamma}, \mathbf{w}, \{\boldsymbol{\omega}_j\}) \ge 0, \quad i \in \mathbb{N}$  (8)

Note that this optimisation problem reduces to a traditional least-squares regression problem if  $n = m_j = 1$  for all j.

# 4 The Intercept Term and Univariate LVR Models

Regression problems often include an intercept term and can be expressed as:

$$\tilde{y} = c + F_{\gamma}(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \dots, \tilde{x}_K) + \epsilon \tag{9}$$

where c is a scalar. The solution for c in a least squares setting is well documented in prior literature, but it is replicated below for completeness.

Defining a shorthand  $\hat{y} \equiv F_{\gamma}(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \dots, \tilde{x}_K)$ , as well as the sample counterparts for  $\tilde{y}$  and  $\hat{y}$  as  $\tilde{y}$  and  $\hat{y}$ , respectively, we can rewrite the optimisation problem from (8) as:

$$\min \|\tilde{\mathbf{y}} - \left(\mathbf{1}_s c + \hat{\tilde{\mathbf{y}}}\right)\|_2^2 = \min \left(\tilde{\mathbf{y}}'\tilde{\mathbf{y}} + \mathbf{1}_s'\mathbf{1}_s c + \hat{\tilde{\mathbf{y}}}'\hat{\tilde{\mathbf{y}}} - 2c\mathbf{1}_s'\tilde{\mathbf{y}} - 2\tilde{\mathbf{y}}'\hat{\tilde{\mathbf{y}}} + 2c\mathbf{1}_s'\hat{\tilde{\mathbf{y}}}\right)$$
(10)

Here, s denotes the sample size of  $\tilde{\mathbf{y}}$  and  $\hat{\mathbf{y}}$ , and  $\mathbf{1}_s$  is a column vector of ones of (sample) length s. Note that this problem is convex with respect to c, so an unconstrained solution can be found by setting the partial derivative to zero:

$$\frac{\partial}{\partial c} \|\tilde{\mathbf{y}} - \left(\mathbf{1}_s c + \hat{\tilde{\mathbf{y}}}\right)\|_2^2 = 2\mathbf{1}_s c - 2\tilde{\mathbf{y}} + 2\hat{\tilde{\mathbf{y}}}$$

$$\mathbf{1}_s c = \tilde{\mathbf{y}} - \hat{\tilde{\mathbf{y}}}$$
(11)

pre-multiplying both sides by  $\mathbf{1}'_s/s$  yields:

$$c = \frac{\mathbf{1}_s'\tilde{\mathbf{y}}}{s} - \frac{\mathbf{1}_s'\tilde{\hat{\mathbf{y}}}}{s} = \overline{\tilde{\mathbf{y}}} - \overline{\hat{\tilde{\mathbf{y}}}}$$
(12)

where  $\overline{\tilde{\mathbf{y}}}$  and  $\overline{\hat{\tilde{\mathbf{y}}}}$  denote the sample means of  $\tilde{\mathbf{y}}$  and  $\hat{\tilde{\mathbf{y}}}$ , respectively. By plugging the unconstrained solution for c back into (10) we arrive at a simplified version of (10):

$$\min \| \left( \tilde{\mathbf{y}} - \mathbf{1}_s \overline{\tilde{\mathbf{y}}} \right) - \left( \hat{\tilde{\mathbf{y}}} - \mathbf{1}_s \overline{\hat{\tilde{\mathbf{y}}}} \right) \|_2^2 = \min \left( \mathbf{\Sigma}_{\tilde{y}} - 2\mathbf{\Sigma}_{\tilde{y}\hat{\tilde{y}}} + \mathbf{\Sigma}_{\hat{\tilde{y}}} \right)$$
(13)

Here,  $\Sigma_A$  denotes the sample covariance matrix over random variable vector A, and  $\Sigma_{AB}$  denotes the sample covariance matrix from A to random variable vector B with elements of A as rows and elements of B as columns. In this case, however,  $\Sigma_{\tilde{y}}$ ,  $\Sigma_{\tilde{y}\hat{y}}$  and  $\Sigma_{\hat{y}}$  all resolve to scalar values because both  $\tilde{y}$  and  $\hat{y}$  are univariate<sup>1</sup>.

Furthermore, in the somewhat trivial case of  $F_{\gamma}(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \dots, \tilde{x}_K) = 0$ , this optimisation problem reduces to:

$$\min \mathbf{\Sigma}_{\tilde{y}} \equiv \min_{\mathbf{w}} \mathbf{w}' \mathbf{\Sigma}_{Y} \mathbf{w} \tag{14}$$

In other words, in their simplest "univariate" form LVR models minimise the *unconditional* variance of a linear combination of variables subject to one or more constraints. This means that problems like PCA and mean-variance portfolio optimisation can be thought of as univariate LVR problems. Conversely, multivariate LVR models can be viewed as advanced portfolio optimisation techniques aimed at minimising the *conditional* variance of an investment strategy with respect to a pre-defined functional relationship (e.g., tracker funds).

<sup>&</sup>lt;sup>1</sup>This simplification is also meaningful if a constraint is imposed on the range of possible solutions for c. The only difference is that if  $c \neq \overline{\hat{\mathbf{y}}} - \overline{\hat{\hat{\mathbf{y}}}}$ , something other than the sample mean would need to be subtracted from  $\tilde{\mathbf{y}}$  and/or  $\hat{\hat{\mathbf{y}}}$  in equation (13), which means that  $\Sigma$  will represent a biased sample variance-covariance estimate for at least one of the variable vectors involved.

### 5 CLARX: A Constrained Latent Variable ARX Model

One of the most popular regression models in economics and finance is the autoregressive model with exogenous inputs (ARX), which expresses the dependent as a linear function of its own past values and the (present and) past values of one or more explanatory variables. Special cases of the ARX model include the autoregressive model (AR) with no exogenous inputs, the lead-lag regression model with no autoregressive element, and multiple linear regression (no lag structure).

This section introduces the LVR counterpart of the ARX model called (C)LARX – a (constrained) latent variable ARX model. As with ARX, the CLARX model comes with a number of interesting special cases, some of which are briefly examined in Section 6.

First, let us define the CLARX model in functional form. Let v be a "version iterator" representing different versions<sup>2</sup> of latent variable j, i.e., a variable identified by LV weight vector  $\boldsymbol{\omega}_j$ . We can denote the total number of versions for variable j as  $V_j \geq 1$ . Let K denote the total number of unique exogenous variables  $\tilde{x}_j$  excluding versions, i.e., the total number of unique LV weight vectors  $\boldsymbol{\omega}_j$ . We can refer to the "autoregressive" versions of the dependent variable as  $\tilde{a}_v \equiv A_v \mathbf{w}$ , with the number of autoregressive versions denoted by the capital  $V_a \geq 0$ . A generic formula for the (C)LARX model can then be written as:

$$\tilde{y} = c + \sum_{v=1}^{V_a} \phi_v \tilde{a}_v + \sum_{v=1}^{V_1} \beta_{1,v} \tilde{x}_{1,v} + \sum_{v=1}^{V_2} \beta_{2,v} \tilde{x}_{2,v} + \dots + \sum_{v=1}^{V_K} \beta_{K,v} \tilde{x}_{K,v} + \epsilon$$

$$\tilde{y} \equiv Y \mathbf{w}, \quad \tilde{a}_v \equiv A_v \mathbf{w}, \quad \tilde{x}_{j,v} \equiv X_{j,v} \boldsymbol{\omega}_j \text{ for } j = 1, \dots, K$$

$$(15)$$

As before, any input variable in this model can be non-latent, as long as the corresponding weight vector only has one non-zero element. The entire problem reduces to the traditional ARX model when all input variables are non-latent.

Models of this kind are more conveniently solved in matrix form. The matrix representation for 15 can be derived with the help of the Kronecker product, traditionally denoted by " $\otimes$ ", and the aforementioned blockwise Kronecker product denoted by " $\odot$ ". The autoregressive terms can be written as:

$$\sum_{v=1}^{V_a} \phi_v \tilde{a}_v = A\left(\boldsymbol{\phi} \otimes \mathbf{w}\right)$$

where  $A = \begin{bmatrix} A_1, & A_2, & \cdots & A_{V_a} \end{bmatrix}$  is a horizontal concatenation of the autoregressive proxy vectors for the dependent, and  $\phi = \begin{bmatrix} \phi_1, & \phi_2, & \cdots & \phi_{V_a} \end{bmatrix}'$  is a column vector containing the respective autoregressive coefficients. Similarly, for each individual explanatory variable  $\tilde{x}_j$  we have:

$$\sum_{v=1}^{V_j} \beta_{j,v} \tilde{x}_{j,v} = X_j \left( \boldsymbol{\beta}_j \otimes \boldsymbol{\omega}_j \right)$$

with  $X_j = \begin{bmatrix} X_{j,1}, & X_{j,2}, & \cdots & X_{j,V_j} \end{bmatrix}$  and  $\beta_j = \begin{bmatrix} \beta_{j,1}, & \beta_{j,2}, & \cdots & \beta_{j,V_j} \end{bmatrix}'$ . All explanatory terms can then be concatenated as follows:

$$\sum_{j=1}^{K} \sum_{v=1}^{V_j} \beta_{j,v} \tilde{x}_{j,v} = \sum_{j=1}^{K} X_j \left( \beta_j \otimes \omega_j \right) = X \left( \beta \odot \omega \right)$$

<sup>&</sup>lt;sup>2</sup>Note that in (C)LARX models, versions may not be synonymous with time series lags. For example,  $\tilde{y} \equiv Y \mathbf{w}$  may represent the return on an investment index and  $\tilde{x}_i \equiv X_i \mathbf{w}$  some other property of the same index such as market capitalization as in the "size" factor of Fama and French (1993).

Here,  $X = \begin{bmatrix} X_1 \mid X_2 \mid \cdots \mid X_K \end{bmatrix}$  is a row vector with K column blocks comprised of the individual  $X_j$ ,  $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1' \mid \boldsymbol{\beta}_2' \mid \cdots \mid \boldsymbol{\beta}_K' \end{bmatrix}'$  is a column vector with K row blocks for the individual  $\boldsymbol{\beta}_j$ , and  $\boldsymbol{\omega}$  has K column blocks containing the individual  $\boldsymbol{\omega}_j$  such that  $\boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{\omega}_1' \mid \boldsymbol{\omega}_2' \mid \cdots \mid \boldsymbol{\omega}_K' \end{bmatrix}'$ .

The complete (C)LARX formula is then concisely defined in matrix form as:

$$Y\mathbf{w} = c + A(\phi \otimes \mathbf{w}) + X(\beta \odot \omega) + \epsilon \tag{16}$$

where Y, A and X together comprise the underlying observed variable space. The least squares optimisation problem for the (C)LARX model can be defined in sample form as:

$$\min_{c, \phi, \beta, \mathbf{w}, \omega} \|\mathbf{Y}\mathbf{w} - \mathbf{1}_{s}c - \mathbf{A} (\phi \otimes \mathbf{w}) - \mathbf{X} (\beta \odot \omega)\|_{2}^{2}$$
(17)

where  $\mathbf{Y}$ ,  $\mathbf{A}$ , and  $\mathbf{X}$  are matrices of sample observations for the random variable vectors Y, A and X, respectively, and s is the length of the sample.

The full CLARX implementation can address the use case of portfolio optimisation by implementing Markowitz-style constraints (e.g., Markowitz (1989)) on the variance and sum of weights for each latent variable<sup>3</sup>. The variance constraint on the dependent takes the form of  $\mathbf{w}' \mathbf{\Sigma}_{Y} \mathbf{w} = \sigma_{y}^{2}$ . The variance constraint on explanatory variable j takes the form of  $\boldsymbol{\omega}'_{j} \mathbf{\Sigma}_{X_{j,c_{j}}} \boldsymbol{\omega}_{j} = \sigma_{j}^{2}$ , where  $c_{j}$  is an arbitrarily chosen version (lag) of  $\tilde{x}_{j}$ . Sum-of-weights constraints take the form of  $\mathbf{1}'_{n}\mathbf{w} = l_{y}$  and  $\mathbf{1}'_{m_{j}}\boldsymbol{\omega}_{j} = l_{j}$ , where  $l_{y}$  and  $l_{j}$  are arbitrary constants. The complete constrained optimization problem then becomes<sup>4</sup>:

$$\min_{c,\phi,\beta,\mathbf{w},\omega} \|\mathbf{Y}\mathbf{w} - \mathbf{1}_{s}c - \mathbf{A} (\boldsymbol{\phi} \otimes \mathbf{w}) - \mathbf{X} (\boldsymbol{\beta} \odot \boldsymbol{\omega}) \|_{2}^{2}$$
s.t. 
$$\mathbf{w}' \boldsymbol{\Sigma}_{Y} \mathbf{w} = \sigma_{y}^{2}, \quad \boldsymbol{\omega}_{j}' \boldsymbol{\Sigma}_{X_{j,c_{j}}} \boldsymbol{\omega}_{j} = \sigma_{j}^{2} \text{ for } 1 \leq j \leq K,$$

$$\mathbf{1}'_{n} \mathbf{w} = l_{y}, \quad \mathbf{1}'_{m_{j}} \boldsymbol{\omega}_{j} = l_{j} \text{ for } 1 \leq j \leq K$$
(18)

The solution for c is covered by section 4 and is given by:

$$c = \overline{\mathbf{Y}}\mathbf{w} - \overline{\mathbf{A}} \left( \boldsymbol{\phi} \otimes \mathbf{w} \right) - \overline{\mathbf{X}} \left( \boldsymbol{\beta} \odot \boldsymbol{\omega} \right)$$
 (19)

Here,  $\overline{\mathbf{Y}}$ ,  $\overline{\mathbf{A}}$  and  $\overline{\mathbf{X}}$  are row vectors containing the column-wise means of  $\mathbf{Y}$ ,  $\mathbf{A}$  and  $\mathbf{X}$ , i.e., the row vectors containing the sample means of the individual variables in Y, A and X. Plugging this solution back into (18) and expanding, we obtain a simplified optimisation problem of the form:

$$\min_{\boldsymbol{\phi}, \boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\omega}} \mathbf{w}' \boldsymbol{\Sigma}_{Y} \mathbf{w} + (\boldsymbol{\phi} \otimes \mathbf{w})' \boldsymbol{\Sigma}_{A} (\boldsymbol{\phi} \otimes \mathbf{w}) + (\boldsymbol{\beta} \odot \boldsymbol{\omega})' \boldsymbol{\Sigma}_{X} (\boldsymbol{\beta} \odot \boldsymbol{\omega}) 
- 2 \left[ \mathbf{w}' \boldsymbol{\Sigma}_{YA} (\boldsymbol{\phi} \otimes \mathbf{w}) + \mathbf{w}' \boldsymbol{\Sigma}_{YX} (\boldsymbol{\beta} \odot \boldsymbol{\omega}) - (\boldsymbol{\phi} \otimes \mathbf{w})' \boldsymbol{\Sigma}_{AX} (\boldsymbol{\beta} \odot \boldsymbol{\omega}) \right] 
\text{s.t. } \mathbf{w}' \boldsymbol{\Sigma}_{Y} \mathbf{w} = \sigma_{y}^{2}, \quad \boldsymbol{\omega}_{j}' \boldsymbol{\Sigma}_{X_{j,c_{j}}} \boldsymbol{\omega}_{j} = \sigma_{j}^{2} \text{ for } 1 \leq j \leq K,$$

$$\mathbf{1}_{n}' \mathbf{w} = l_{y}, \quad \mathbf{1}_{m_{j}}' \boldsymbol{\omega}_{j} = l_{j} \text{ for } 1 \leq j \leq K$$

 $<sup>^3</sup>$ Section 6 examines a counterpart of this model with minimal constraints.

<sup>&</sup>lt;sup>4</sup>To the author's best knowledge, the closest precedent for this optimisation problem in prior literature is the EDACCA model defined in Xu and Zhu (2024). The EDACCA model has a number of important differences, including a different set of constraints and the use of a single explanatory variable.

This is a convex problem with equality constraints, which means it can be solved using the method of Lagrange multipliers (LM). Like with the rest of the model, we can represent the LM terms in matrix form with the help of the following vectors:

$$\boldsymbol{\vartheta}_{x} = \begin{bmatrix} \sigma_{x,1}^{2}, & \sigma_{x,2}^{2}, & \sigma_{x,3}^{2}, & \dots & \sigma_{x,K}^{2} \end{bmatrix}', \qquad \boldsymbol{\lambda}_{x} = \begin{bmatrix} \lambda_{x,1}, & \lambda_{x,2}, & \lambda_{x,3}, & \dots & \lambda_{x,K} \end{bmatrix}'$$

$$\boldsymbol{l}_{p} = \begin{bmatrix} l_{p,1}, & l_{p,2}, & l_{p,3}, & \dots & l_{p,K} \end{bmatrix}', \qquad \boldsymbol{\lambda}_{p} = \begin{bmatrix} \lambda_{p,1}, & \lambda_{p,2}, & \lambda_{p,3}, & \dots & \lambda_{p,K} \end{bmatrix}'$$

 $\vartheta_x$  is a  $K \times 1$  column vector of variance targets for the chosen versions of  $\tilde{x}_j$ ;  $\boldsymbol{l}_p$  is a  $K \times 1$  column vector of sum-of-weights targets for the respective LV weight vectors  $\omega_j$ ; and  $\lambda_x$  and  $\lambda_p$  are the corresponding vectors of LM coefficients. The full constraint terms for the Lagrangian function can then be defined in matrix form with the help of the blockwise direct sum operator introduced in Section 2. The sum-of-weights constraints become:

$$\left(oldsymbol{\omega}^{\oplus}
ight)' \mathbf{1}_{oldsymbol{\omega}} = oldsymbol{l}_p$$

where  $\mathbf{1}_{\omega}$  is a column vector of ones with the same row block structure as  $\omega$ .

For the variance constraints we require an indexer identifying which version of each  $\tilde{x}_j$  has a variance constraint assigned to it. Let  $\langle \mathbf{u}_j \rangle$  be a sequence of K logical vectors, each of size  $V_j$  (i.e., the number of versions of  $\tilde{x}_j$ ) with a value of 1 in position of  $c_j$ , i.e., the version of  $\tilde{x}_j$  that has a variance constraint, and zeros elsewhere. Column vector  $\mathbf{u} \equiv [\langle \mathbf{u}_j \rangle]$  is a vertical concatenation of  $\langle \mathbf{u}_j \rangle$  which has the same size and row block structure as the vector  $\boldsymbol{\beta}$ . We also need a block-diagonal matrix  $\boldsymbol{\Sigma}_X^d$  containing the covariance matrices of the individual  $X_{j,v_j}$  along the diagonal, but no covariances across different variables or versions. Once again, this can be done with the help of the blockwise direct sum operator:

$$oldsymbol{\Sigma}_{X}^{d} = \left[ \left( \mathbf{X} - \overline{\mathbf{X}} 
ight)^{\oplus} 
ight]' \left( \mathbf{X} - \overline{\mathbf{X}} 
ight)^{\oplus} = egin{pmatrix} oldsymbol{\Sigma}_{X_{1}}^{d}, & \mathbf{0}, & \cdots & \mathbf{0} \ \mathbf{0}, & oldsymbol{\Sigma}_{X_{2}}^{d}, & \cdots & \mathbf{0} \ dots & dots & \ddots & dots \ \mathbf{0}, & \mathbf{0}, & \cdots & oldsymbol{\Sigma}_{X_{K}}^{d} \end{pmatrix}$$

where

$$oldsymbol{\Sigma}_{X_j}^d = \left[ \left( \mathbf{X}_j - \overline{\mathbf{X}_j} 
ight)^\oplus 
ight]' \left( \mathbf{X}_j - \overline{\mathbf{X}_j} 
ight)^\oplus = egin{pmatrix} oldsymbol{\Sigma}_{X_{j,1}}, & \mathbf{0}, & \cdots & \mathbf{0} \ \mathbf{0}, & oldsymbol{\Sigma}_{X_{j,2}}, & \cdots & \mathbf{0} \ dots & dots & \ddots & dots \ \mathbf{0}, & \mathbf{0}, & \cdots & oldsymbol{\Sigma}_{X_{j,V_s}} \end{pmatrix}$$

for  $1 \le j \le K$ 

The variance constraints on all  $\tilde{x}_j$  can then be expressed using a single quadratic form:

$$\left[\left(\mathbf{u}\odotoldsymbol{\omega}
ight)'
ight]^{\oplus}oldsymbol{\Sigma}_{X}^{d}\left(\mathbf{u}\odotoldsymbol{\omega}
ight)=oldsymbol{artheta}_{x}$$

The full Lagrangian function for (20) can then be written as:

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\omega}, \lambda_{y}, \lambda_{l}, \boldsymbol{\lambda}_{x}, \boldsymbol{\lambda}_{p}) = \mathbf{w}' \boldsymbol{\Sigma}_{Y} \mathbf{w} + (\boldsymbol{\phi} \otimes \mathbf{w})' \boldsymbol{\Sigma}_{A} (\boldsymbol{\phi} \otimes \mathbf{w}) + (\boldsymbol{\beta} \odot \boldsymbol{\omega})' \boldsymbol{\Sigma}_{X} (\boldsymbol{\beta} \odot \boldsymbol{\omega})$$

$$- 2 \mathbf{w}' \boldsymbol{\Sigma}_{YA} (\boldsymbol{\phi} \otimes \mathbf{w}) - 2 \mathbf{w}' \boldsymbol{\Sigma}_{YX} (\boldsymbol{\beta} \odot \boldsymbol{\omega}) + 2 (\boldsymbol{\phi} \otimes \mathbf{w})' \boldsymbol{\Sigma}_{AX} (\boldsymbol{\beta} \odot \boldsymbol{\omega})$$

$$+ \lambda_{y} (\mathbf{w}' \boldsymbol{\Sigma}_{Y} \mathbf{w} - \sigma_{y}^{2}) + \lambda_{l} (\mathbf{1}'_{n} \mathbf{w} - l_{y}) + \boldsymbol{\lambda}'_{x} \left\{ \left[ (\mathbf{u} \odot \boldsymbol{\omega})' \right]^{\oplus} \boldsymbol{\Sigma}_{X}^{d} (\mathbf{u} \odot \boldsymbol{\omega}) - \boldsymbol{\vartheta}_{x} \right\}$$

$$+ \boldsymbol{\lambda}'_{p} \left[ (\boldsymbol{\omega}^{\oplus})' \mathbf{1}_{\boldsymbol{\omega}} - \boldsymbol{l}_{p} \right]$$

$$(21)$$

Note that in expanded form the terms under  $\lambda_x$  and  $\lambda_p$  resolve to:

$$\begin{split} & \boldsymbol{\lambda}_{x}^{\prime} \left\{ \left[ \left( \mathbf{u} \odot \boldsymbol{\omega} \right)^{\prime} \right]^{\oplus} \boldsymbol{\Sigma}_{X}^{d} \left( \mathbf{u} \odot \boldsymbol{\omega} \right) - \boldsymbol{\vartheta}_{x} \right\} = \sum_{j=1}^{K} \lambda_{x,j} \left( \boldsymbol{\omega}_{j}^{\prime} \boldsymbol{\Sigma}_{X_{j,c_{j}}} \boldsymbol{\omega}_{j} - \sigma_{x,j}^{2} \right) \\ & \boldsymbol{\lambda}_{p}^{\prime} \left[ \left( \boldsymbol{\omega}^{\oplus} \right)^{\prime} \mathbf{1}_{\boldsymbol{\omega}} - \boldsymbol{l}_{p} \right] = \sum_{j=1}^{K} \lambda_{p,j} \left( \boldsymbol{\omega}_{j}^{\prime} \mathbf{1}_{m_{j}} - l_{p,j} \right) \end{split}$$

where  $m_j$  represents the number of observed variables used to approximate  $\tilde{x}_j$ .

The solution for (21) is obtained by setting various partial derivatives to zero.

First, we note that the properties of the Kronecker product and the blockwise Kronecker product allow us to factorise  $(\phi \otimes \mathbf{w})$  and  $(\beta \odot \omega)$  for compatibility with traditional matrix calculus:

$$(\phi \otimes \mathbf{w}) = (I_{V_n} \otimes \mathbf{w}) \, \phi = (\phi \otimes I_n) \, \mathbf{w} \tag{22a}$$

$$(\boldsymbol{\beta} \odot \boldsymbol{\omega}) = (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega}) \, \boldsymbol{\beta} = (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}}) \, \boldsymbol{\omega}$$
 (22b)

Here,  $I_a$  represents an identity matrix of size a (where a is a scalar value), while  $I_a$  represents an identity matrix with the same size and row block structure as vector  $\mathbf{a}$  (vectors are conventionally represented by bold letters). The property of the Kronecker product relevant for (22a) is well established but reiterated for completeness in equations (43-44). The factorisation of the blockwise Kronecker product used for (22b) is proved in Appendix B. Taking  $\Sigma_{BA}$  to denote the transpose of  $\Sigma_{AB}$ , the solution for  $\beta$  becomes:

$$\frac{\partial}{\partial \boldsymbol{\beta}} \mathcal{L} = 2 \left[ (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \boldsymbol{\Sigma}_{X} (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega}) \boldsymbol{\beta} - (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \boldsymbol{\Sigma}_{XY} \mathbf{w} + (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \boldsymbol{\Sigma}_{XA} (\boldsymbol{\phi} \otimes \mathbf{w}) \right]$$
$$\boldsymbol{\beta} = \left[ (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \boldsymbol{\Sigma}_{X} (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega}) \right]^{-1} \left[ (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \boldsymbol{\Sigma}_{XY} \mathbf{w} - (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \boldsymbol{\Sigma}_{XA} (\boldsymbol{\phi} \otimes \mathbf{w}) \right]$$
(23)

Similarly, recalling that  $(\phi \otimes \mathbf{w}) = (I_{V_a} \otimes \mathbf{w}) \phi$ , the solution for  $\phi$  is:

$$\frac{\partial}{\partial \boldsymbol{\phi}} \mathcal{L} = 2 \left[ (I_{V_a} \otimes \mathbf{w})' \boldsymbol{\Sigma}_A (I_{V_a} \otimes \mathbf{w}) \boldsymbol{\phi} - (I_{V_a} \otimes \mathbf{w})' \boldsymbol{\Sigma}_{AY} \mathbf{w} + (I_{V_a} \otimes \mathbf{w})' \boldsymbol{\Sigma}_{AX} (\boldsymbol{\beta} \odot \boldsymbol{\omega}) \right]$$

$$\boldsymbol{\phi} = \left[ (I_{V_a} \otimes \mathbf{w})' \boldsymbol{\Sigma}_A (I_{V_a} \otimes \mathbf{w}) \right]^{-1} \left[ (I_{V_a} \otimes \mathbf{w})' \boldsymbol{\Sigma}_{AY} \mathbf{w} - (I_{V_a} \otimes \mathbf{w})' \boldsymbol{\Sigma}_{AX} (\boldsymbol{\beta} \odot \boldsymbol{\omega}) \right]$$
(24)

For the dependent weight vector  $\mathbf{w}$ , we recall that  $(\phi \otimes \mathbf{w}) = (\phi \otimes I_n) \mathbf{w}$ . The partial derivative of  $\mathcal{L}$  with respect to  $\mathbf{w}$  is then given by:

$$\frac{\partial}{\partial \mathbf{w}} \mathcal{L} = 2 \left[ (1 + \lambda_y) \mathbf{\Sigma}_Y + (\phi \otimes I_n)' \mathbf{\Sigma}_A (\phi \otimes I_n) - (\phi \otimes I_n)' \mathbf{\Sigma}_{AY} - \mathbf{\Sigma}_{YA} (\phi \otimes I_n) \right] \mathbf{w}$$
$$- 2 \left[ \mathbf{\Sigma}_{YX} (\boldsymbol{\beta} \odot \boldsymbol{\omega}) - (\phi \otimes I_n)' \mathbf{\Sigma}_{AX} (\boldsymbol{\beta} \odot \boldsymbol{\omega}) \right] + \lambda_l \mathbf{1}_n$$

Note that  $\frac{\partial}{\partial \mathbf{w}} \mathbf{w}' \mathbf{\Sigma}_{YA} (\boldsymbol{\phi} \otimes I_n) \mathbf{w}$  resolves to  $[(\boldsymbol{\phi} \otimes I_n)' \mathbf{\Sigma}_{AY} + \mathbf{\Sigma}_{YA} (\boldsymbol{\phi} \otimes I_n)] \mathbf{w}$  because the quadratic form is not symmetric. Setting the partial derivative to zero and expressing in terms of  $\mathbf{w}$ , we get:

$$(1 + \lambda_{y}) \Sigma_{Y} \mathbf{w} = \left[ (\phi \otimes I_{n})' \Sigma_{AY} + \Sigma_{YA} (\phi \otimes I_{n}) - (\phi \otimes I_{n})' \Sigma_{A} (\phi \otimes I_{n}) \right] \mathbf{w}$$

$$+ \left[ \Sigma_{YX} - (\phi \otimes I_{n})' \Sigma_{AX} \right] (\beta \odot \omega) - \frac{\lambda_{l}}{2} \mathbf{1}_{n}$$
(25)

For ease of notation, define:

$$\mathbf{v}_{1} = \left[ \left( \boldsymbol{\phi} \otimes I_{n} \right)' \boldsymbol{\Sigma}_{AY} + \boldsymbol{\Sigma}_{YA} \left( \boldsymbol{\phi} \otimes I_{n} \right) - \left( \boldsymbol{\phi} \otimes I_{n} \right)' \boldsymbol{\Sigma}_{A} \left( \boldsymbol{\phi} \otimes I_{n} \right) \right] \mathbf{w}$$

$$\mathbf{v}_{2} = \left[\mathbf{\Sigma}_{YX} - \left(\mathbf{\phi} \otimes I_{n}\right)' \mathbf{\Sigma}_{AX}\right] \left(\mathbf{\beta} \odot \mathbf{\omega}\right)$$

Equation (25) then becomes:

$$(1 + \lambda_y) \Sigma_Y \mathbf{w} = \mathbf{v}_1 + \mathbf{v}_2 - \frac{\lambda_l}{2} \mathbf{1}_n$$
 (26)

Setting  $\rho_y = (1 + \lambda_y)$ ,  $\rho_l = \frac{\lambda_l}{2}$  and pre-multiplying both sides by  $\frac{1}{\rho_y} \Sigma_Y^{-1}$  we get:

$$\mathbf{w} = \frac{1}{\rho_u} \left[ \mathbf{\Sigma}_Y^{-1} \left( \mathbf{v}_1 + \mathbf{v}_2 \right) - \rho_l \mathbf{\Sigma}_Y^{-1} \mathbf{1}_n \right]$$
 (27)

For the weight vector  $\boldsymbol{\omega}$ , we recall that  $(\boldsymbol{\beta} \odot \boldsymbol{\omega}) = (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}}) \boldsymbol{\omega}$ . Furthermore, we can refactor the constraint terms for compatibility with traditional matrix calculus using the properties of the blockwise direct sum operator. For the portfolio constraints on  $\boldsymbol{\omega}_j$  we can apply Proposition 2.3 to obtain:

$$\left(\omega^{\oplus}
ight)' \mathbf{1}_{oldsymbol{\omega}} = \left(\mathbf{1}_{oldsymbol{\omega}}^{\oplus}
ight)' oldsymbol{\omega}$$

For the variance constraints on  $\tilde{x}_j$ , we note three things. First of all, because **u** has the same length and row block structure as  $\beta$ , the term  $(\mathbf{u} \odot \boldsymbol{\omega})$  can be factorised in the same way as  $(\beta \odot \boldsymbol{\omega})$ , namely:

$$(\mathbf{u} \odot \boldsymbol{\omega}) = (\mathbf{u} \odot I_{\boldsymbol{\omega}}) \boldsymbol{\omega} = (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega}) \mathbf{u}$$

Second, by applying Proposition 2.1 we have:

$$\left[ (\mathbf{u} \odot \boldsymbol{\omega})' 
ight]^{\oplus} = \left[ (\mathbf{u} \odot \boldsymbol{\omega})^{\oplus} 
ight]'$$

Third, the operation  $(\mathbf{u} \odot \boldsymbol{\omega})$  can be expressed as a left-multiplication over the sequence of blocks in  $\boldsymbol{\omega}$ , i.e.,  $(\mathbf{u} \odot \boldsymbol{\omega}) = (\mathbf{u} \odot I_{\boldsymbol{\omega}}) \boldsymbol{\omega} = \left[\left\langle \left(\mathbf{u}_j \otimes \left\langle I_{\boldsymbol{\omega}} \right\rangle_j \right) \boldsymbol{\omega}_j | 1 \leq j \leq K \right\rangle \right]_v$ , which means that according to Proposition A1 we have:

$$(\mathbf{u} \odot \boldsymbol{\omega})^{\oplus} \equiv [(\mathbf{u} \odot I_{\boldsymbol{\omega}}) \, \boldsymbol{\omega}]^{\oplus} = (\mathbf{u} \odot I_{\boldsymbol{\omega}}) \, \boldsymbol{\omega}^{\oplus}$$

Putting these transformations together, we can rewrite the blockwise quadratic form as:

$$\begin{bmatrix} (\mathbf{u} \odot \boldsymbol{\omega})' \end{bmatrix}^{\oplus} \boldsymbol{\Sigma}_{X}^{d} (\mathbf{u} \odot \boldsymbol{\omega}) = \begin{bmatrix} (\mathbf{u} \odot \boldsymbol{\omega})^{\oplus} \end{bmatrix}' \boldsymbol{\Sigma}_{X}^{d} (\mathbf{u} \odot \boldsymbol{\omega}) = \{ [(\mathbf{u} \odot I_{\boldsymbol{\omega}}) \boldsymbol{\omega}]^{\oplus} \}' \boldsymbol{\Sigma}_{X}^{d} (\mathbf{u} \odot \boldsymbol{\omega}) \\ = [(\mathbf{u} \odot I_{\boldsymbol{\omega}}) \boldsymbol{\omega}^{\oplus}]' \boldsymbol{\Sigma}_{X}^{d} (\mathbf{u} \odot \boldsymbol{\omega}) = (\boldsymbol{\omega}^{\oplus})' (\mathbf{u} \odot I_{\boldsymbol{\omega}})' \boldsymbol{\Sigma}_{x}^{d} (\mathbf{u} \odot I_{\boldsymbol{\omega}}) \boldsymbol{\omega} \end{bmatrix}$$

Applying these transformations, the partial derivative of (21) with respect to  $\omega$  is:

$$\frac{\partial}{\partial \boldsymbol{\omega}} \mathcal{L} = 2 \left( \boldsymbol{\beta} \odot I_{\boldsymbol{\omega}} \right)' \left[ \boldsymbol{\Sigma}_{X} \left( \boldsymbol{\beta} \odot I_{\boldsymbol{\omega}} \right) \boldsymbol{\omega} - \boldsymbol{\Sigma}_{XY} \mathbf{w} + \boldsymbol{\Sigma}_{XA} \left( \boldsymbol{\phi} \otimes \mathbf{w} \right) \right] + 2 \mathbf{M}_{2} \boldsymbol{\omega}^{\oplus} \boldsymbol{\lambda}_{x} + \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \boldsymbol{\lambda}_{p}$$

where  $\mathbf{M}_2 = (\mathbf{u} \odot I_{\boldsymbol{\omega}})' \boldsymbol{\Sigma}_X^d (\mathbf{u} \odot I_{\boldsymbol{\omega}})$ . Furthermore, Propositions 2.2 and B1 prove that:

$$oldsymbol{\omega}^{\oplus}oldsymbol{\lambda}_x = (oldsymbol{\omega}\odot I_K)\,oldsymbol{\lambda}_x = (oldsymbol{\lambda}_x\odot I_{oldsymbol{\omega}})\,oldsymbol{\omega}$$

Re-arranging for  $\omega$ , we get:

$$\boldsymbol{\omega} = \left[ \left( \boldsymbol{\beta} \odot I_{\boldsymbol{\omega}} \right)' \boldsymbol{\Sigma}_{X} \left( \boldsymbol{\beta} \odot I_{\boldsymbol{\omega}} \right) + \mathbf{M}_{2} \left( \boldsymbol{\lambda}_{x} \odot I_{\boldsymbol{\omega}} \right) \right]^{-1} \left[ \mathbf{v}_{3} - \frac{1}{2} \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \boldsymbol{\lambda}_{p} \right]$$
(28)

with  $\mathbf{v}_3 = (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}})' [\boldsymbol{\Sigma}_{XY} - \boldsymbol{\Sigma}_{XA} (\boldsymbol{\phi} \otimes I_n)] \mathbf{w}$ . The derivations for the Lagrange multipliers are a bit more involved and deferred to Appendices C and D. Finally, the full solution for the CLARX problem can be expressed as the following fixed point problem:

$$\mathbf{w} = \frac{1}{\rho_y} \left[ \mathbf{\Sigma}_Y^{-1} \left( \mathbf{v}_1 + \mathbf{v}_2 \right) - \rho_l \mathbf{\Sigma}_Y^{-1} \mathbf{1}_n \right]$$
(29a)

$$\boldsymbol{\omega} = \left[ (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}})' \boldsymbol{\Sigma}_{X} (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}}) + \mathbf{M}_{2} (\boldsymbol{\lambda}_{x} \odot I_{\boldsymbol{\omega}}) \right]^{-1} \left[ \mathbf{v}_{3} - \frac{1}{2} \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \boldsymbol{\lambda}_{p} \right]$$
(29b)

$$\phi = \left[ (I_{V_a} \otimes \mathbf{w})' \, \mathbf{\Sigma}_A \, (I_{V_a} \otimes \mathbf{w}) \right]^{-1} \left( I_{V_a} \otimes \mathbf{w} \right)' \left[ \mathbf{\Sigma}_{AY} \mathbf{w} - \mathbf{\Sigma}_{AX} \, (\boldsymbol{\beta} \odot \boldsymbol{\omega}) \right]$$
(29c)

$$\boldsymbol{\beta} = \left[ (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \boldsymbol{\Sigma}_{X} (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega}) \right]^{-1} (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \left[ \boldsymbol{\Sigma}_{XY} \mathbf{w} - \boldsymbol{\Sigma}_{XA} (\boldsymbol{\phi} \otimes \mathbf{w}) \right]$$
(29d)

$$\begin{cases}
\mathbf{w} = \frac{1}{\rho_{y}} \left[ \mathbf{\Sigma}_{Y}^{-1} \left( \mathbf{v}_{1} + \mathbf{v}_{2} \right) - \rho_{l} \mathbf{\Sigma}_{Y}^{-1} \mathbf{1}_{n} \right] & (29a) \\
\boldsymbol{\omega} = \left[ \left( \boldsymbol{\beta} \odot I_{\boldsymbol{\omega}} \right)' \mathbf{\Sigma}_{X} \left( \boldsymbol{\beta} \odot I_{\boldsymbol{\omega}} \right) + \mathbf{M}_{2} \left( \boldsymbol{\lambda}_{x} \odot I_{\boldsymbol{\omega}} \right) \right]^{-1} \left[ \mathbf{v}_{3} - \frac{1}{2} \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \boldsymbol{\lambda}_{p} \right] & (29b) \\
\boldsymbol{\phi} = \left[ \left( I_{V_{a}} \otimes \mathbf{w} \right)' \mathbf{\Sigma}_{X} \left( I_{V_{a}} \otimes \mathbf{w} \right) \right]^{-1} \left( I_{V_{a}} \otimes \mathbf{w} \right)' \left[ \mathbf{\Sigma}_{AY} \mathbf{w} - \mathbf{\Sigma}_{AX} \left( \boldsymbol{\beta} \odot \boldsymbol{\omega} \right) \right] & (29c) \\
\boldsymbol{\beta} = \left[ \left( I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega} \right)' \mathbf{\Sigma}_{X} \left( I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega} \right) \right]^{-1} \left( I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega} \right)' \left[ \mathbf{\Sigma}_{XY} \mathbf{w} - \mathbf{\Sigma}_{XA} \left( \boldsymbol{\phi} \otimes \mathbf{w} \right) \right] & (29d) \\
\boldsymbol{\rho}_{y} = \frac{\left( n \mathbf{w} - l_{y} \mathbf{1}_{n} \right)' \left( \mathbf{v}_{1} + \mathbf{v}_{2} \right)}{n \sigma_{y}^{2} - l_{y} \mathbf{1}_{n}' \mathbf{\Sigma}_{Y} \mathbf{w}} & (29e) \\
\boldsymbol{\rho}_{l} = \frac{1}{n} \left[ \mathbf{1}_{n}' \left( \mathbf{v}_{1} + \mathbf{v}_{2} \right) - \rho_{y} \mathbf{1}_{n}' \mathbf{\Sigma}_{Y} \mathbf{w} \right] & (29f) \\
\boldsymbol{\lambda}_{x} = \left[ \mathbf{M}_{1} \mathbf{\Theta} - \mathbf{L} \left( \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \right)' \mathbf{M}_{2} \boldsymbol{\omega}^{\oplus} \right]^{-1} \left( \boldsymbol{\omega}^{\oplus} \mathbf{M}_{1} - \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \mathbf{L} \right)' \left( \mathbf{v}_{3} - \mathbf{v}_{4} \right) & (29g) \\
\boldsymbol{\lambda}_{p} = 2 \mathbf{M}_{1}^{-1} \left( \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \right)' \left[ \left( \mathbf{v}_{3} - \mathbf{v}_{4} \right) - \mathbf{M}_{2} \boldsymbol{\omega}^{\oplus} \boldsymbol{\lambda}_{x} \right] & (29h)
\end{cases}$$

$$\rho_{l} = \frac{1}{n} \left[ \mathbf{1}'_{n} \left( \mathbf{v}_{1} + \mathbf{v}_{2} \right) - \rho_{y} \, \mathbf{1}'_{n} \boldsymbol{\Sigma}_{Y} \mathbf{w} \right]$$
(29f)

$$\lambda_{x} = \left[ \mathbf{M}_{1} \mathbf{\Theta} - \mathbf{L} \left( \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \right)' \mathbf{M}_{2} \boldsymbol{\omega}^{\oplus} \right]^{-1} \left( \boldsymbol{\omega}^{\oplus} \mathbf{M}_{1} - \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \mathbf{L} \right)' (\mathbf{v}_{3} - \mathbf{v}_{4})$$
(29g)

$$\lambda_p = 2\mathbf{M}_1^{-1} \left(\mathbf{1}_{\boldsymbol{\omega}}^{\oplus}\right)' \left[ (\mathbf{v}_3 - \mathbf{v}_4) - \mathbf{M}_2 \boldsymbol{\omega}^{\oplus} \lambda_x \right]$$
(29h)

with the following shorthand notations:

 $\mathbf{1}_{\omega}$  a column vector of ones with the same length and block structure as  $\boldsymbol{\omega}$ 

$$\mathbf{v}_{1} = \left[ (\boldsymbol{\phi} \otimes I_{n})' \, \boldsymbol{\Sigma}_{AY} + \boldsymbol{\Sigma}_{YA} \, (\boldsymbol{\phi} \otimes I_{n}) - (\boldsymbol{\phi} \otimes I_{n})' \, \boldsymbol{\Sigma}_{A} \, (\boldsymbol{\phi} \otimes I_{n}) \right] \mathbf{w}$$

$$\mathbf{v}_{2} = \left[ \boldsymbol{\Sigma}_{YX} - (\boldsymbol{\phi} \otimes I_{n})' \, \boldsymbol{\Sigma}_{AX} \right] (\boldsymbol{\beta} \odot \boldsymbol{\omega})$$

$$\mathbf{v}_{3} = (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}})' \left[ \boldsymbol{\Sigma}_{XY} - \boldsymbol{\Sigma}_{XA} \, (\boldsymbol{\phi} \otimes I_{n}) \right] \mathbf{w}$$

$$\mathbf{v}_{4} = (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}})' \, \boldsymbol{\Sigma}_{X} \, (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}}) \, \boldsymbol{\omega}$$

$$\boldsymbol{\Theta} = \operatorname{diag}(\boldsymbol{\vartheta}_{x})$$

$$\mathbf{L} = \operatorname{diag}(\boldsymbol{l}_p)$$

$$\mathbf{M}_1 = \left(\mathbf{1}_{m{\omega}}^{\oplus}\right)' \mathbf{1}_{m{\omega}}^{\oplus}$$

$$\mathbf{M}_{2} = (\mathbf{u} \odot I_{\boldsymbol{\omega}})' \, \boldsymbol{\Sigma}_{X}^{d} \, (\mathbf{u} \odot I_{\boldsymbol{\omega}})$$

This problem can be solved using fixed point iteration with initial guesses required for  $\mathbf{w}$ ,  $\boldsymbol{\omega}$ , and either  $\phi$  or  $\beta$ . Equations can be estimated in the same order as shown above. With initial guesses for  $\mathbf{w}, \boldsymbol{\omega}$  and  $\boldsymbol{\phi}$ , the first iteration can start at (29d). Four matrices need to be inverted at each iteration step and cannot become singular. That said, equations (29c) and (29d) can be reformulated in terms of the Moore-Penrose inverses (Penrose (1955); Bjerhammar (1951); Moore (1920)) of the matrices  $(\mathbf{A} - \overline{\mathbf{A}}) (I_{V_a} \otimes \mathbf{w})$  and  $(\mathbf{X} - \overline{\mathbf{X}}) (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})$ , respectively, which makes them solvable by SVD.

#### Special cases of (C)LARX 6

The CLARX model has a number of interesting special cases which arise under various simplifying assumptions. This section briefly introduces four such models:

- 1. LARX: A CLARX model with minimal constraints
- 2. LSR: An LVR equivalent of a univariate lead-lag regression
- 3. LVMR: An LVR equivalent of a multiple linear regression
- 4. LAR: A latent variable autoregressive model (no exogenous inputs)

#### 6.1LARX: CLARX without the C(onstraints)

(C)LARX models with a latent dependent variable require a constraint on the variance of  $\tilde{y}$  to eliminate the trivial solution given by  $\mathbf{w} = \mathbf{0}$ . All other constraints are optional with minor caveats. Setting the optional Lagrange multiplier terms from (29) to zero produces:

$$\mathbf{w} = \frac{1}{\rho_y} \mathbf{\Sigma}_Y^{-1} \left( \mathbf{v}_1 + \mathbf{v}_2 \right)$$
(30a)

$$\boldsymbol{\omega} = \left[ (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}})' \boldsymbol{\Sigma}_{X} (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}}) \right]^{-1} (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}})' \left[ \boldsymbol{\Sigma}_{XY} - \boldsymbol{\Sigma}_{XA} (\boldsymbol{\phi} \otimes I_{n}) \right] \mathbf{w}$$
(30b)

$$\phi = \left[ (I_{V_a} \otimes \mathbf{w})' \, \mathbf{\Sigma}_A \, (I_{V_a} \otimes \mathbf{w}) \right]^{-1} (I_{V_a} \otimes \mathbf{w})' \left[ \mathbf{\Sigma}_{AY} \mathbf{w} - \mathbf{\Sigma}_{AX} \, (\boldsymbol{\beta} \odot \boldsymbol{\omega}) \right]$$
(30c)

$$\boldsymbol{\beta} = \left[ (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \, \boldsymbol{\Sigma}_{X} \, (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega}) \right]^{-1} \, (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \, \left[ \boldsymbol{\Sigma}_{XY} - \boldsymbol{\Sigma}_{XA} \, (\boldsymbol{\phi} \otimes I_{n}) \right] \mathbf{w}$$
(30d)

$$\begin{cases} \mathbf{w} = \frac{1}{\rho_{y}} \mathbf{\Sigma}_{Y}^{-1} (\mathbf{v}_{1} + \mathbf{v}_{2}) \\ \boldsymbol{\omega} = \left[ (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}})' \, \mathbf{\Sigma}_{X} (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}}) \right]^{-1} (\boldsymbol{\beta} \odot I_{\boldsymbol{\omega}})' \left[ \mathbf{\Sigma}_{XY} - \mathbf{\Sigma}_{XA} (\boldsymbol{\phi} \otimes I_{n}) \right] \mathbf{w} \\ \boldsymbol{\phi} = \left[ (I_{V_{a}} \otimes \mathbf{w})' \, \mathbf{\Sigma}_{A} (I_{V_{a}} \otimes \mathbf{w}) \right]^{-1} (I_{V_{a}} \otimes \mathbf{w})' \left[ \mathbf{\Sigma}_{AY} \mathbf{w} - \mathbf{\Sigma}_{AX} (\boldsymbol{\beta} \odot \boldsymbol{\omega}) \right] \\ \boldsymbol{\beta} = \left[ (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \, \mathbf{\Sigma}_{X} (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega}) \right]^{-1} (I_{\boldsymbol{\beta}} \odot \boldsymbol{\omega})' \left[ \mathbf{\Sigma}_{XY} - \mathbf{\Sigma}_{XA} (\boldsymbol{\phi} \otimes I_{n}) \right] \mathbf{w} \\ \boldsymbol{\rho}_{y} = \frac{\mathbf{w}' (\mathbf{v}_{1} + \mathbf{v}_{2})}{\sigma_{y}^{2}} \end{cases}$$
(30a)

with:

$$\mathbf{v}_{1} = \left[ (\boldsymbol{\phi} \otimes I_{n})' \, \boldsymbol{\Sigma}_{AY} + \boldsymbol{\Sigma}_{YA} \, (\boldsymbol{\phi} \otimes I_{n}) - (\boldsymbol{\phi} \otimes I_{n})' \, \boldsymbol{\Sigma}_{A} \, (\boldsymbol{\phi} \otimes I_{n}) \right] \mathbf{w}$$

$$\mathbf{v}_{2} = \left[ \boldsymbol{\Sigma}_{YX} - (\boldsymbol{\phi} \otimes I_{n})' \, \boldsymbol{\Sigma}_{AX} \right] (\boldsymbol{\beta} \odot \boldsymbol{\omega})$$

This specification offers a better intuition about the meaning of the individual coefficient vectors. As constraints are removed, the solution for the respective vector reduces to an OLS formula conditional on the values of the other vectors. This solution does, however, come with at least two caveats: First of all, for any estimate of  $\tilde{y}$  given by the LV weight vector  $\hat{\mathbf{w}}$ , an equally valid estimate is given by  $-\hat{\mathbf{w}}$ . Second for each pair of the estimated  $\hat{\omega}_j$  and  $\hat{\beta}_j$ , an equally valid estimate is given by  $k\hat{\omega}_j$  and  $\frac{1}{k}\hat{\beta}_j$ where k is an arbitrary non-zero constant<sup>5</sup>. This can, however, be rectified by rescaling the relevant vectors after the fact and/or by modifying the fixed point algorithm to ensure that either  $\omega_j$  or  $\beta_j$  is always normalised (e.g., by enforcing  $\mathbf{1}'_{V_j}\boldsymbol{\beta}_j=1$  at each iteration step).

#### 6.2 (C)LSR: A Parsimonious Lead-Lag Regression

A key difference between (C)LARX and traditional ARX models lies in how regression coefficients are mapped to the observed variables. ARX models always assign a unique response coefficient to each variable, whereas (C)LARX models allow some coefficients to enter the equation more than once. In other words, (C)LARX models will often be more parsimonious than OLS models defined over the same observed variable space.

 $<sup>{}^{5}\</sup>mathbf{A}\otimes\mathbf{B}=(k\mathbf{A})\otimes(\frac{1}{k}\mathbf{B})$  for any two matrices  $\mathbf{A},\mathbf{B}$  and non-zero constant k by the properties of the Kronecker product.

The difference is best exemplified by a class of models in which an observed dependent variable y is a function of V versions of a single latent explanatory variable  $\tilde{x}$  with m proxies. In a time series context, the version iterator v becomes a lag iterator  $\tau$  such that:

$$y_{t} = c + \sum_{\tau=1}^{F} \beta_{\tau} \tilde{x}_{t-\tau} + \epsilon_{t}$$

$$\tilde{x}_{t} = X_{t} \boldsymbol{\omega}$$
(31)

We can call this a (constrained) Latent Shock Regression model, or (C)LSR for short. The LSR equation can be written in matrix form as:

$$y = c + X \left( \beta \otimes \omega \right) + \epsilon \tag{32}$$

Here, X has dimensions  $1 \times m$ ,  $\omega$  has dimensions  $m \times 1$ , and  $\beta$  has dimensions  $F \times 1$ . The unconstrained solution to this problem is given by:

$$\begin{cases} \boldsymbol{\omega} = \left[ (\boldsymbol{\beta} \otimes I_m)' \, \boldsymbol{\Sigma}_X \, (\boldsymbol{\beta} \otimes I_m) \right]^{-1} \, (\boldsymbol{\beta} \otimes I_m)' \, \boldsymbol{\Sigma}_{Xy} \\ \boldsymbol{\beta} = \left[ (I_F \otimes \boldsymbol{\omega})' \, \boldsymbol{\Sigma}_X \, (I_F \otimes \boldsymbol{\omega}) \right]^{-1} \, (I_F \otimes \boldsymbol{\omega})' \, \boldsymbol{\Sigma}_{Xy} \end{cases}$$
(33a)

$$c = \overline{\mathbf{y}} - \overline{\mathbf{X}} \left( \boldsymbol{\beta} \otimes \boldsymbol{\omega} \right) \tag{34}$$

This model has m observed explanatory variables with F lags each. A traditional lead-lag regression would require mF response coefficients to estimate the relationship, while LSR only requires F + mcoefficients: F for the vector  $\boldsymbol{\beta}$  and m for the vector  $\boldsymbol{\omega}$ . If F is equal to m, the difference becomes  $m^2$  vs 2m – a substantial reduction in complexity for large values of m. This parsimony is achieved by means of a simplifying assumption: In LSR models, all explanatory variables affect the dependent with a shared lag profile: the lag profile of  $\tilde{x}$ , given by  $\beta$ . The weights of the explanatory variables in  $\tilde{x}$  are time-invariant and given by  $\omega$ . The resonance coefficient for lag  $\tau$  of observed variable i is then given by the product of the i'th element of  $\omega$  and the  $\tau$ 'th element of  $\beta$ .

#### (C)LVMR: Reducing (C)LARX to CCA and OLS 6.3

Another interesting subcategory of (C)LARX models is one in which all variables are either non-latent or enter the equation exactly once. We can broadly categorise these models as (Constrained) Latent Variable Multiple Regression (LVMR) models because of their resemblance to the class of models examined in Burnham et al. (1996). These models have two simplifying features compared to the full CLARX specification: First of all, there are no autoregressive lags and hence no vector  $\phi$ . Second, because each explanatory variable j only has one lag, there is only one element in each  $\beta_i$ , so the entire vector  $\boldsymbol{\beta}$  can be "scaled away" as long as there are no scaling constraints on  $\boldsymbol{\omega}$ . As a result, the regression formula reduces to:

$$Y\mathbf{w} = c + X\boldsymbol{\omega} + \epsilon \tag{35}$$

The solution to this problem is given by:

$$\begin{cases} \mathbf{w} = \frac{1}{\rho_y} \mathbf{\Sigma}_Y^{-1} \mathbf{\Sigma}_{YX} \boldsymbol{\omega} \\ \boldsymbol{\omega} = \mathbf{\Sigma}_X^{-1} \mathbf{\Sigma}_{XY} \mathbf{w} \\ \rho_y = \frac{\mathbf{w}' \mathbf{\Sigma}_{YX} \boldsymbol{\omega}}{\sigma_z^2} \end{cases}$$
(36a)

$$\omega = \Sigma_X^{-1} \Sigma_{XY} \mathbf{w} \tag{36b}$$

$$\rho_y = \frac{\mathbf{w}' \, \mathbf{\Sigma}_{YX} \boldsymbol{\omega}}{\sigma_y^2} \tag{36c}$$

$$c = \overline{\mathbf{Y}}\mathbf{w} - \overline{\mathbf{X}}\boldsymbol{\omega} \tag{37}$$

LVMR models sit at the cusp between (C)LARX, CCA and traditional least squares regression. First of all, equation (36b) represents the least squares solution for a linear regression of  $\tilde{y}$  on the individual observed variables in X. Semantically, the vector  $\omega$  could just as well be called  $\beta$ , and whether any blocks of  $\omega$  represent LV weight vectors for some latent variable(s)  $\tilde{x}_i$  is a question of interpretation only. As a corollary, LVMR models reduce to standard multiple regression models when  $\tilde{y}$  is nonlatent. Furthermore, solving (36) is equivalent to finding the dominant canonical variates for Y and X as observed by Dong and Qin (2018). Substituting (36b) into (36a) and setting  $\sigma_y^2 = 1$ , we obtain:

$$\mathbf{w} = \frac{\mathbf{\Sigma}_{Y}^{-1} \mathbf{\Sigma}_{YX} \mathbf{\Sigma}_{X}^{-1} \mathbf{\Sigma}_{XY} \mathbf{w}}{\mathbf{w}' \mathbf{\Sigma}_{Y}^{-1} \mathbf{\Sigma}_{YX} \mathbf{\Sigma}_{X}^{-1} \mathbf{\Sigma}_{XY} \mathbf{w}}$$
(38)

which is the mathematical formula for canonical correlation analysis.

#### 6.4 (C)LAR: (Constrained) Latent Variable Autoregressive Models

Just like the ARX model subsumes the autoregressive model (AR) as a special case, the (C)LARX model subsumes a latent variable autoregressive model which we can refer to as (C)LAR. This class of models has been covered relatively well by papers from other disciplines. For example, Dong and Qin (2018) considers a similar family of models under the name DiCCA (dynamic inner CCA), while Qin (2021) proposes a LaVAR (latent vector autoregression) algorithm for achieving a full canonical decomposition of the latent autoregressive structure in Y allowing for interactions. Furthermore, first-order LAR models bear a strong resemblance to Min/Max Autocorrelation Factors (MAF) first introduced in Switzer and Green (1984) and since popularised in the geosciences, although a more detailed comparison with these models left to future research.

(C)LAR models can be defined by stripping away the exogenous term from equation (16), which results in the following formula:

$$Y\mathbf{w} = c + A\left(\phi \otimes \mathbf{w}\right) + \epsilon \tag{39}$$

The main use case of this class of models outside of economics of finance is the decomposition of multivariate sensor data into time-persistent signals on the one hand, and serially uncorellated white noise on the other. In the context of investment management, the same concept can be applied to derive trend following investment strategies such as price momentum. If Y contains the returns on assets in a fund's investment opportunity set and Y**w** represents the return on an investment strategy characterised by capital allocation weights w, then the strategy that has the strongest price momentum or reversal signal (i.e., maximum absolute correlation between past returns and future returns) is given by:

$$\mathbf{w} = \frac{1}{\rho_y} \mathbf{\Sigma}_Y^{-1} \left( \mathbf{v}_1 - \rho_l \mathbf{1}_n \right)$$
 (40a)

$$\boldsymbol{\phi} = \left[ (I_{V_a} \otimes \mathbf{w})' \, \boldsymbol{\Sigma}_A \, (I_{V_a} \otimes \mathbf{w}) \right]^{-1} (I_{V_a} \otimes \mathbf{w})' \, \boldsymbol{\Sigma}_{AY} \mathbf{w}$$
(40b)

$$\begin{cases} \mathbf{w} = \frac{1}{\rho_y} \mathbf{\Sigma}_Y^{-1} (\mathbf{v}_1 - \rho_l \mathbf{1}_n) \\ \phi = \left[ (I_{V_a} \otimes \mathbf{w})' \mathbf{\Sigma}_A (I_{V_a} \otimes \mathbf{w}) \right]^{-1} (I_{V_a} \otimes \mathbf{w})' \mathbf{\Sigma}_{AY} \mathbf{w} \end{cases}$$

$$\begin{cases} \rho_y = \frac{(n\mathbf{w} - l_y \mathbf{1}_n)' \mathbf{v}_1}{n\sigma_y^2 - l_y \mathbf{1}_n' \mathbf{\Sigma}_Y \mathbf{w}} \end{cases}$$

$$(40a)$$

$$\begin{cases} \rho_l = \frac{1}{n} \left[ \mathbf{1}_n' \mathbf{v}_1 - \rho_y \mathbf{1}_n' \mathbf{\Sigma}_Y \mathbf{w} \right] \end{cases}$$

$$(40a)$$

$$\rho_l = \frac{1}{n} \left[ \mathbf{1}'_n \mathbf{v}_1 - \rho_y \, \mathbf{1}'_n \mathbf{\Sigma}_Y \mathbf{w} \right] \tag{40d}$$

with 
$$\mathbf{v}_{1} = \left[ \left( \boldsymbol{\phi} \otimes I_{n} \right)' \boldsymbol{\Sigma}_{AY} + \boldsymbol{\Sigma}_{YA} \left( \boldsymbol{\phi} \otimes I_{n} \right) - \left( \boldsymbol{\phi} \otimes I_{n} \right)' \boldsymbol{\Sigma}_{A} \left( \boldsymbol{\phi} \otimes I_{n} \right) \right] \mathbf{w}$$

For a strategy with zero-sum weights, the solution for  $\rho_y$  further reduces to  $\rho_y = \frac{\mathbf{w}'\mathbf{v}_1}{\sigma_y^2}$ .

Let us also briefly consider the simplest type of LAR problem, which is a first-order autoregressive model for a variance-covariance stationary  $\tilde{y}$  of the form  $\tilde{y}_t = c + \phi \tilde{y}_{t-1} + \epsilon$  with  $\tilde{y}_t = Y \mathbf{w}$  and  $\tilde{y}_{t-1} = A \mathbf{w}$ . In this case the solution for  $\mathbf{w}$  further reduces to:

$$\phi \mathbf{w} = \frac{1}{2} \left( \mathbf{\Sigma}_A^{-1} \mathbf{\Sigma}_{AY} + \mathbf{\Sigma}_Y^{-1} \mathbf{\Sigma}_{YA} \right) \mathbf{w}$$
 (41)

Here,  $\Sigma_A^{-1}\Sigma_{AY}$  and  $\Sigma_Y^{-1}\Sigma_{YA}$  are matrices of least squares regression coefficients, e.g., the first column of  $\Sigma_A^{-1}\Sigma_{AY}$  contains the least squares coefficients from a regression of the first variable in Y on all variables in A. Each vector  $\mathbf{w}_i$  which solves (41) is an eigenvector of  $(\Sigma_A^{-1}\Sigma_{AY} + \Sigma_Y^{-1}\Sigma_{YA})$ .  $2\phi$  is the matching eigenvalue, wherein  $\phi$  is the first-order autocorrelation coefficient of  $\tilde{y}$ . In other words, the dominant eigenvector  $\mathbf{w}_1$  produces the strongest autocorrelation in  $\tilde{y}$  in absolute terms.

In an investment management context, (41) is a formula for constructing momentum and reversal strategies based on first-order autocorrelation. Let Y and A represent asset returns for an investment opportunity set at times t and t-1, respectively. The capital allocation weights for each strategy are the eigenvectors of the matrix  $(\Sigma_A^{-1}\Sigma_{AY} + \Sigma_Y^{-1}\Sigma_{YA})$ . The strategy with the strongest (weakest) signal is given by the first (last) eigenvector. The direction of the signal (i.e., momentum vs reversal) is determined by the sign of the corresponding eigenvalue (i.e., the sign of the autocorrelation coefficient).

# 7 Empirical Application: Stock markets and economic activity in the US

As a simple example of how (C)LARX models can be used in the real world, let us examine the relationship between equity market performance and real economic activity in the United States. A good starting point for this analysis is provided by Ball and French (2021) who find that de-trended levels of the S&P 500 index have in-sample predictive power over the de-trended levels of real US GDP. The best-performing model is found to be one with lags zero to three of the S&P 500 as well as two autoregressive terms, reaching an adjusted R-squared of 66.61% for a quarterly data sample between Q1 1999 and Q4 2020.

The theoretical foundation for this relationship is relatively simple: Stock prices reflect discounted expectations of future earnings, and earnings are related to economic activity. In practice, however, market aggregates and macroeconomic aggregates are designed to measure different things. Companies in the S&P 500 are weighted based on their market capitalisation rather than their relative contributions to the real economy, while US GDP is designed to measure *all* economic activity in the US – not just that of the Fortune 500 companies. It stands to reason that the true strength of the relationship between stock performance and real economic output would be underestimated by a model linking the S&P 500 to US GDP.

With the (C)LARX methodology, equity market performance and real economic output can both be viewed as latent variables and optimised to conform to the chosen functional relationship. The latent measure of equity market performance effectively becomes a tracker index for real growth expectations in large-cap US stocks, while the latent measure of real economic output is optimised to reflect the real economic output of the Fortune 500 companies rather than the whole economy.

The latent measures can be constructed from the same building blocks as their non-latent counterparts, in order to gain additional insights about the compositional differences between the S&P 500 and US GDP. For the latent equity measure, we can use ten S&P 500 GICS level 1 sector sub-indices<sup>6</sup>; for the

<sup>&</sup>lt;sup>6</sup>As of 2016, listed real estate (RE) was added as the eleventh GICS level 1 sector of the S&P 500. The RE sector is excluded from this study for two reasons: First of all, its data history only starts in Q4 2001 and would reduce the sample size from 138 quarterly observations to 90. Second, the RE sector only has a 2.25% weight in the S&P 500 as of April 2025 – the second lowest weight after materials at 1.99%.

real economic output measure, we can use the five GDP expenditure components. This should reveal whether the relationship between the S&P 500 and real US GDP is distorted by differences in sector composition, the relative importance of various sources of demand, or both. In principle, this study can be performed at higher levels of granularity, such as industries or even single stocks for the equity measure and/or itemised national accounts for the real output measure. However, such a detailed analysis is left to future research.

### 7.1 Data and Methodology

The model identified in Ball and French (2021) is used as the basis for this study with three noteworthy changes. First of all, performance is measured out of sample using rolling regressions. Exponentially decaying sample weights with a half-life of 10 years are used to capture changes in the relationship over time.

Second, percent changes are used instead of de-trended levels. Log-returns are calculated for the S&P 500 and its sector constituents. Annualised log-percent changes are calculated for US GDP and its expenditure components. Revised estimates are used for all economic aggregates following Ball and French (2021)'s tentative finding that the link is stronger between equity performance and revised GDP numbers as opposed to point-in-time ("vintage") releases.

Third, the sample period covered by this paper is Q4 1989 to Q1 2025, capturing the full available data history for the relevant S&P 500 sectors at the time of writing, with the COVID lockdown period of Q2 and Q3 2020 removed as a statistical outlier. US GDP shows a contraction of 8.2% in Q2 2020 followed by a 7.5% rebound in Q3 2020 – a -13.8 sigma event and a 12.6 sigma event, respectively, based on the standard deviation of quarterly US GDP growth excluding these two quarters.

A total of four regression models are estimated:

Baseline OLS/ARX model: Real GDP growth g vs S&P 500 returns r:

$$g_t = c + \sum_{\tau=1}^{2} \phi_{t-\tau} g_{t-\tau} + \sum_{\tau=0}^{3} \beta_{t-\tau} r_{t-\tau} + \epsilon$$
 (42)

LARX model a): Real GDP growth g vs a Latent measure of market growth expectations,  $\tilde{r}$ :

$$g_t = c + \sum_{\tau=1}^{2} \phi_{t-\tau} g_{t-\tau} + \sum_{\tau=0}^{3} \beta_{t-\tau} \tilde{r}_{t-\tau} + \epsilon$$
 (42a)

LARX model b): A latent measure of the real economic output of Fortune 500 companies,  $\tilde{g}$ , vs S&P 500 returns r:

$$\tilde{g}_t = c + \sum_{\tau=1}^2 \phi_{t-\tau} \tilde{g}_{t-\tau} + \sum_{\tau=0}^3 \beta_{t-\tau} r_{t-\tau} + \epsilon$$
 (42b)

LARX model c): A latent measure of the real economic output of Fortune 500 companies,  $\tilde{g}$ , vs a Latent measure of market growth expectations,  $\tilde{r}$ :

$$\tilde{g}_t = c + \sum_{\tau=1}^2 \phi_{t-\tau} \tilde{g}_{t-\tau} + \sum_{\tau=0}^3 \beta_{t-\tau} \tilde{r}_{t-\tau} + \epsilon$$
 (42c)

No constraints are imposed on the latent variables except for the necessary variance constraint on  $\tilde{g}$  in equations (42b) and (42c). The variance target for  $\tilde{g}$  is reverse-engineered to ensure that the expenditure weights add up to 1 as in the official GDP number.

A minimum of 40 degrees of freedom is set as a requirement for producing a forecast. This corredponds to 10 years of quarterly data on top of one data point for each estimated coefficient including Lagrange multipliers. An additional three data points are lost to the lag operator and one to the percent change

calculation. As a result, forecast coverage starts in Q3 2002 for the baseline model (longest) and in Q3 2006 for the model with the all latent variables (shortest).

Historical data for US GDP and its expenditure components are retrieved from the Economic Database of the Federal Reserve Bank of St. Louis ("FRED"). Historical index levels for the S&P 500 and its GICS level 1 sector constituents are retrieved from Investing.com. A full data reference can be found in Table 1.

### 7.2 Out-of-Sample Forecasting Performance

Figure 1 plots the rolling out-of-sample (OOS) predictions of the four regression models for the full available forecast coverage. Each plot overlays the actual values of the dependent, as well as a naïve forecast for the dependent based on a rolling sample mean (used as the benchmark). The grey text boxes report each model's Mean Squared Prediction Error (MSPE) as a percentage of the MSPE of the rolling sample mean model, which can be viewed as a type of out-of-sample R-squared metric.

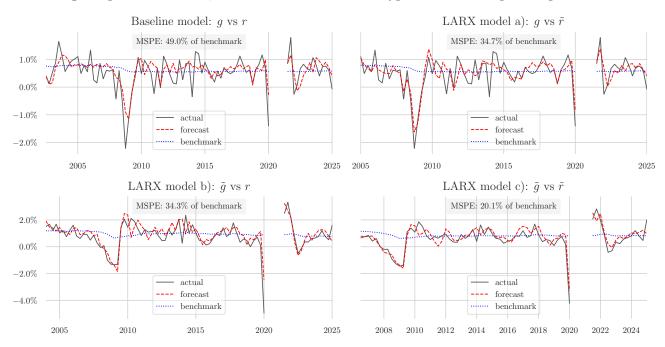


Figure 1: Out-of-sample forecasting performance of regression models (42)-(42c)

The baseline model from Ball and French (2021) (top left) does well despite the design changes implemented by this paper. Its MSPE is 51% lower than that of the naïve forecast – reasonably close to the reported in-sample adjusted R-squared of 66.61%. This suggests that the baseline model is well specified.

The models with latent variables further improve on these results. A LARX model with a latent measure of market growth expectations (top right) produces an MSPE 65.3% below benchmark. A LARX model with a latent measure of the real economic output of the S&P 500 (bottom left) prodices an MSPE 65.7% below benchmark. A LARX model with latent measures for both economic output and market growth expectations (bottom right) performs best with an MSPE 79.9% below benchmark.

#### 7.3 Insights from the Latent Variable Weight Vectors

The relative accuracy of the out-of-sample forecasts suggests that the latent measures capture new information about the relationship between stock returns and economic growth. The two latent variables contribute equally to the improvement in MSPE, which drops from around 50% to 35% of benchmark in both cases. The model with two latent variables yields twice the improvement of the single-LV specifications, which suggests that the two latent measures capture complementary information. For a deeper understanding of what drives these results, we can look at the LV component

weights of the two latent measures as estimated by LARX model c), which shows the strongest outof-sample performance.

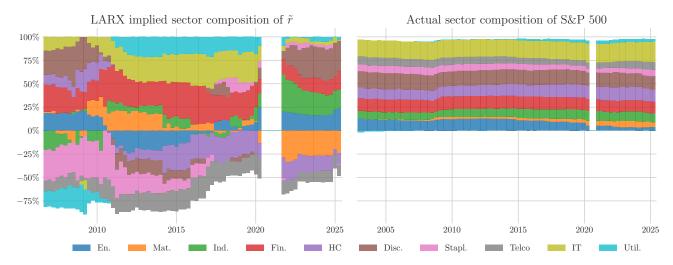


Figure 2: Sector composition: Latent Measure of Market Growth Expectations vs the S&P 500

Figure 2 plots the evolution of the sector weights in the latent measure of market growth expectations (left) against the actual sector weights of the S&P 500 (right). Positive weights in the LV weight vector are scaled to 100%. Sector weights of the S&P 500 are approximated using a rolling regression of S&P 500 returns on the coincident returns of the GICS level 1 sector indices.

At least two straightforward observations can be made here. First of all, sector rotations are important for gauging growth expectations in the equity market. The latent measure assigns negative weights to various sectors throughout the study period, with between 50% and 80% of its positive sector weights counter-balanced by negative weights in other sectors. This is not something that a long only, capital weighted index like the S&P 500 is designed to capture.

Second, the composition of the latent equity measure seems to fluctuate strongly over time. It is unlikely that all of these fluctuations have straightforward economic interpretations, but some of them may be related to changes in the relative value added to the economy by different sectors, while others may be explained by structural trends or events. For example, the healthcare sector ("HC") has a positive weight in the latent measure until around 2010, which marks the introduction of the Affordable Care Act ("Obamacare"), and a negative weight thereafter.

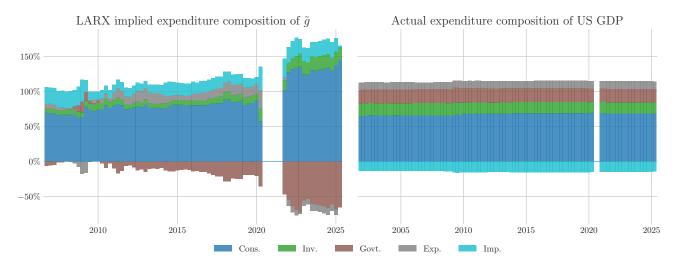


Figure 3: Expenditure composition: Latent Economic Activity Measure vs US GDP

Figure 3 plots the evolution of the expenditure weights in the latent measure of the real economic output of S&P 500 companies (left) against the evolution of the expenditure weights in the official

measure of real US GDP (right).

Somewhat expectedly, consumer spending ("Cons.") is the most relevant component for both the US stock market and US GDP. The role of private investment ("Inv.") is largely similar in the two measures.

At the same time, there is a strong difference in the role of government spending (Govt.). In the latent measure, stronger government spending, all else being equal, tends to follow after periods of weaker equity market performance and vice versa, outside of the immediate aftermath of the 2008 crisis. There are at least two alternative interpretations for this result: On the one hand, the US fiscal authorities might view equity market strength as a sign of a stronger economy and pare back discretionary spending in response; On the other hand, equity investors might view government spending as a positive in crisis periods and a negative otherwise. Either way, both interpretations support the Kaynesian argument for government spending as a counter-cyclical buffer, and the effect has become more pronounced after the COVID pandemic, perhaps owing to a sharp rise in the cost of US government debt since 2021.

Lastly, international trade plays an important and complex role in both the real economy and the equity market. Reinbold and Wen (2019) offers a brief primer on the history of the US trade balance across the value chain at different stages of industrialisation. From an accounting perspective, imports (Imp.) subtract from GDP while exports (Exp.) add to it. However, in modern times US goods are often produced abroad, which means that they need to be imported for domestic consumption but don't need to be exported to be sold elsewhere. This should create a weaker link between stock returns and exports on the one hand, and a positive link between stock returns and imports on the other. The (C)LARX model is able to capture this to some degree by assigning a positive weight to imports and a time-varying weight to exports, in contrast to the official GDP methodology.

# 8 Concluding Remarks

This paper proposes a new latent variable regression (LVR) framework for finance and economics which can be viewed as an off-shoot of Canonical Correlation Analysis. A fixed point solution is derived for a family of linear LVR models called (C)LARX – a constrained LVR implementation of the traditional ARX methodology. A minor contribution is also made to the field of matrix calculus: A new blockwise direct sum operator is introduced and applied to solve a class of Lagrangian optimisation problems with piecemeal constraints on the target coefficient vector.

In a stylised empirical application, (C)LARX models are used to examine how well the stock market predicts real economic activity in the United States. By "synthesizing" latent measures for equity market performance and real economic output, (C)LARX models outperform the baseline OLS specification out of sample and offer novel analytical insights about the relationship, including the importance of sector rotations in gauging investor growth expectations, as well as the complex interplay between stock market returns and concepts like government spending and international trade.

(C)LARX models have many more potential use cases in economics and finance. Although few variables in our field are traditionally thought of as being latent, their estimation methodologies may not always be fit for purpose in a given research context, and their measurement accuracy can sometimes be called into question. Consider, for example, diffusion indices of business activity (e.g., Owens and Sarte (2005)), surveys of consumer sentiment (e.g., Curtin et al. (2000)), composite indicators of financial stress (e.g., Hollo et al. (2012)), or various rules of thumb in investment management such as holding a 60/40 allocation of stocks and bonds in lieu of a mean-variance efficient portfolio (e.g., Ambachtsheer (1987)) or "buying past winners and selling past losers" as a means of capturing asset price momentum (e.g., Jegadeesh and Titman (1993)). (C)LARX and other LVR models can be used to improve the accuracy of our approximations for these variables based on economic theories describing the relationships between them.

Much room is left for future methodological research as well. One important topic not covered in this

paper is that of statistical significance (e.g., see Bagozzi et al. (1981)) and feature selection. Furthermore, (C)LARX models can be modified in much the same way as traditional ARX models, including moving average errors and/or conditional heteroskedasticity via maximum likelihood estimation, various forms of coefficient regularisation such as LASSO, Ridge or Elastic Net (e.g., see Vinod (1976)), and various covariance adjustment techniques such as Generalised Least Squares (e.g., Aitken (1936)) and portfolio-style covariance shrinkage (e.g., Ledoit and Wolf (2020)). Non-linear LVR models can also be developed to address more complex relationships and concepts, such as functional dependencies involving volatility and valuation ratios.

To conclude, LVR models like (C)LARX can be viewed as a rather natural extension of traditional regression analysis. At the same time, an effective application of these models requires a slight paradigm shift on the part of the researcher in interpreting of both theory and data. This may pose a conceptual challenge, but it also creates a wide range of opportunities for future empirical and methodological work.

#### References

- James H. Stock and Francesco Trebbi. Retrospectives: Who invented instrumental variable regression? Journal of Economic Perspectives, 17(3):177–194, 2003.
- Leonard E. Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563, 1966. ISSN 00034851. URL http://www.jstor.org/stable/2238772.
- H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936. ISSN 00063444. URL http://www.jstor.org/stable/2333955.
- Herman Wold. Soft modelling: the basic design and some extensions. Systems under indirect observation, Part II, pages 36–37, 1982.
- Herman Wold. Soft modelling by latent variables: The non-linear iterative partial least squares (nipals) approach. *Journal of Applied Probability*, 12(S1):117–142, 1975. doi: 10.1017/S0021900200047604.
- Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11):559–572, 1901. doi: 10.1080/14786440109462720. URL https://doi.org/10.1080/14786440109462720.
- H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, "September" 1933. ISSN 0022-0663. doi: 10.1037/h0071325. URL http://dx.doi.org/10.1037/h0071325.
- Jushan Bai and Serena Ng. Evaluating latent and observed factors in macroeconomics and finance. Journal of Econometrics, 131(1):507–537, 2006. ISSN 0304-4076. doi: 10.1016/j.jeconom.2005.01. 015. URL https://www.sciencedirect.com/science/article/pii/S0304407605000205.
- Seung C Ahn, Stephan Dieckmann, and M Fabricio Perez. Exploring common factors in the term structure of credit spreads: the use of canonical correlations. *Available at SSRN 984245*, 2012. doi: http://dx.doi.org/10.2139/ssrn.984245. URL https://papers.ssrn.com/sol3/papers.cfm? abstract\_id=984245.
- H. D. Vinod. Canonical ridge and econometrics of joint production. *Journal of Econometrics*, 4 (2):147–166, 1976. ISSN 0304-4076. doi: 10.1016/0304-4076(76)90010-5. URL https://www.sciencedirect.com/science/article/pii/0304407676900105.
- H. D. Vinod. Econometrics of joint production. *Econometrica*, 36(2):322–336, 1968. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1907492.

- Alison J. Burnham, Roman Viveros, and John F. MacGregor. Frameworks for latent variable multivariate regression. *Journal of Chemometrics*, 10(1):31–45, 1996. doi: 10.1002/(SICI)1099-128X(199601)10:1\(\alpha\)31::AID-CEM398\(\alpha\)3.0.CO;2-1. URL https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291099-128X%28199601%2910%3A1%3C31%3A%3AAID-CEM398%3E3.0.CO%3B2-1.
- Alison J. Burnham, Roman Viveros, and John F. MacGregor. Latent variable multivariate regression modeling. *Chemometrics and Intelligent Laboratory Systems*, 48(2):167–180, 1999. ISSN 0169-7439. doi: 10.1016/S0169-7439(99)00018-0. URL https://www.sciencedirect.com/science/article/pii/S0169743999000180.
- Li Wang, Lei-hong Zhang, Zhaojun Bai, and Ren-Cang Li. Orthogonal canonical correlation analysis and applications. *Optimization Methods and Software*, 35(4):787–807, 2020. doi: 10.1080/10556788. 2019.1700257. URL https://doi.org/10.1080/10556788.2019.1700257.
- Liang Dai, Guodong Du, Jia Zhang, Candong Li, Rong Wei, and Shaozi Li. Joint multilabel classification and feature selection based on deep canonical correlation analysis. *Concurrency and Computation: Practice and Experience*, 32(22):e5864, 2020. doi: 10.1002/cpe.5864. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5864.
- Steven Van Vaerenbergh, Ignacio Santamaria, Victor Elvira, and Matteo Salvatori. Pattern localization in time series through signal-to-model alignment in latent space. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2711–2715, 2018. doi: 10.1109/ICASSP.2018.8461890.
- Eric C. Chi, Genevera I. Allen, Hua Zhou, Omid Kohannim, Kenneth Lange, and Paul M. Thompson. Imaging genetics via sparse canonical correlation analysis. In 2013 IEEE 10th International Symposium on Biomedical Imaging, pages 740–743, "April" 2013. doi: 10.1109/ISBI.2013.6556581. URL https://ieeexplore.ieee.org/abstract/document/6556581.
- Viivi Uurtio, João M. Monteiro, Jaz Kandola, John Shawe-Taylor, Delmiro Fernandez-Reyes, and Juho Rousu. A tutorial on canonical correlation methods. *ACM Comput. Surv.*, 50(6), "November" 2017. ISSN 0360-0300. doi: 10.1145/3136624. URL https://doi.org/10.1145/3136624.
- Yining Dong and S. Joe Qin. Dynamic latent variable analytics for process operations and control. Computers & Chemical Engineering, 114:69-80, 2018. ISSN 0098-1354. doi: 10. 1016/j.compchemeng.2017.10.029. URL https://www.sciencedirect.com/science/article/pii/S0098135417303848. FOCAPO/CPC 2017.
- S. Joe Qin. Latent vector autoregressive modeling for reduced dimensional dynamic feature extraction and prediction. In 2021 60th IEEE Conference on Decision and Control (CDC), pages 3689–3694, 2021. doi: 10.1109/CDC45484.2021.9683255.
- C. G. Khatri and C. Radhakrishna Rao. Solutions to some functional equations and their applications to characterization of probability distributions. Sankhyā: The Indian Journal of Statistics, Series A (1961-2002), 30(2):167–180, 1968. ISSN 0581572X. URL http://www.jstor.org/stable/25049527.
- Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1):3-56, 1993. ISSN 0304-405X. doi: 10.1016/0304-405X(93) 90023-5. URL https://www.sciencedirect.com/science/article/pii/0304405X93900235.
- Harry M. Markowitz. *Mean—Variance Analysis*, pages 194–198. Palgrave Macmillan UK, London, 1989. ISBN 978-1-349-20213-3. doi: 10.1007/978-1-349-20213-3\_21. URL https://doi.org/10.1007/978-1-349-20213-3\_21.
- Bo Xu and Qinqin Zhu. An efficient dynamic auto-regressive cca for time series imputation with irregular sampling. *IEEE Transactions on Automation Science and Engineering*, 21(1):442–451, 2024. doi: 10.1109/TASE.2022.3217451.

- R. Penrose. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406–413, 1955. doi: 10.1017/S0305004100030401.
- Arne Bjerhammar. Application of calculus of matrices to method of least squares: with special reference to geodetic calculations. Kungl. Tekniska högskolans handlingar = Transactions of the Royal Institute of Technology, Stockholm, Sweden. Elanders Boktryckeri Aktiebolag, 1951. URL https://cir.nii.ac.jp/crid/1130000798238614528.
- E. H. Moore. On the reciprocal of the general algebraic matrix. Bulletin of the American Mathematical Society, 26:294–295, 1920. URL https://cir.nii.ac.jp/crid/1573387450082342272.
- P Switzer and A Green. Min/max autocorrelation factors for multivariate spatial imagery. Technical report, Dept. of Statistics, Stanford University, 1984.
- Christopher Ball and Jack French. Exploring what stock markets tell us about gdp in theory and practice. Research in Economics, 75(4):330–344, 2021. ISSN 1090-9443. doi: https://doi.org/10.1016/j.rie.2021.09.002. URL https://www.sciencedirect.com/science/article/pii/S1090944321000399.
- Brian Reinbold and Yi Wen. Historical u.s. trade deficits. *Economic Synopses*, 2019(13), 2019. doi: 10.20955/es.2019.13.
- Raymond E. Owens and Pierre-Daniel Sarte. How well do diffusion indexes capture business cycles? a spectral analysis. FRB Richmond Economic Quarterly, 91(4):23–42, 2005. URL https://ssrn.com/abstract=2185567.
- Richard Curtin, Stanley Presser, and Eleanor Singer. The effects of response rate changes on the index of consumer sentiment\*. *Public Opinion Quarterly*, 64(4):413–428, 02 2000. ISSN 0033-362X. doi: 10.1086/318638. URL https://doi.org/10.1086/318638.
- Daniel Hollo, Manfred Kremer, and Marco Lo Duca. Ciss a composite indicator of systemic stress in the financial system. *ECB Working Paper*, (1426), 2012. doi: 10.2139/ssrn.2018792. URL https://ssrn.com/abstract=2018792.
- Keith P. Ambachtsheer. Pension fund asset allocation: In defense of a 60/40 equity/debt asset mix. Financial Analysts Journal, 43(5):14-24, 1987. doi: 10.2469/faj.v43.n5.14. URL https://doi.org/10.2469/faj.v43.n5.14.
- Narasimhan Jegadeesh and Sheridan Titman. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance*, 48(1):65–91, "March" 1993. URL https://ideas.repec.org/a/bla/jfinan/v48y1993i1p65-91.html.
- Richard P. Bagozzi, Claes Fornell, and David F. Larcker. Canonical correlation analysis as a special case of a structural relations model. *Multivariate Behavioral Research*, 16(4):437–454, 1981. doi: 10.1207/s15327906mbr1604\\_2. PMID: 26812673.
- A. C. Aitken. Iv.—on least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1936. doi: 10.1017/S0370164600014346.
- Olivier Ledoit and Michael Wolf. The power of (non-)linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*, 20(1):187–218, 06 2020. ISSN 1479-8409. doi: 10.1093/jjfinec/nbaa007. URL https://doi.org/10.1093/jjfinec/nbaa007.

# A Commutativity of the blockwise direct sum operator

**Proposition A1.** Let S be a set of all matrix sequences of length k, and let the sequence  $\langle \mathbf{A} \rangle \equiv \langle \mathbf{A}_i | 1 \leq i \leq k \rangle$  be an element in S. Let the matrix  $\mathbf{A}^{\oplus}$  represent the direct sum over the elements in  $\langle \mathbf{A} \rangle$ . For any function  $f: S \to S$ , if  $f(\langle \mathbf{A} \rangle)$  can be expressed as a sequence  $\langle \mathbf{M}_i \mathbf{A}_i \rangle \equiv \langle \mathbf{M}_i \mathbf{A}_i | 1 \leq i \leq k \rangle$ 

for some arbitrary sequence of matrices  $\langle \mathbf{M}_i | 1 \leq i \leq k \rangle$ , then  $f(\langle \mathbf{A} \rangle)_v^{\oplus} = [f(\langle \mathbf{A}_v^{\oplus} \rangle)]_v$ . If  $f(\langle \mathbf{A} \rangle)_v^{\oplus} = [f(\langle \mathbf{A}_v^{\oplus} \rangle)]_v$ . If  $f(\langle \mathbf{A} \rangle)_v^{\oplus} = [f(\langle \mathbf{A}_v^{\oplus} \rangle)]_v$ . If  $f(\langle \mathbf{A} \rangle)_v^{\oplus} = [f(\langle \mathbf{A}_v^{\oplus} \rangle)]_v$ .

*Proof.* For a block matrix  $\mathbf{A}$ , take  $\langle \mathbf{A} \rangle$  to denote the sequence of the blocks in  $\mathbf{A}$ . For a sequence of matrices  $\langle \mathbf{A} \rangle$ , denote its *i*'th element by  $\langle \mathbf{A} \rangle_i$ . For the case of left-multiplication we then have:

$$f\left(\langle \mathbf{A} \rangle\right)_{v}^{\oplus} = \langle \mathbf{M}_{i} \mathbf{A}_{i} \rangle_{v}^{\oplus} = \left( \begin{array}{c|ccc} \mathbf{M}_{1} \mathbf{A}_{1}, & \mathbf{0}, & \cdots & \mathbf{0} \\ \hline \mathbf{0}, & \mathbf{M}_{2} \mathbf{A}_{2}, & \cdots & \mathbf{0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{0}, & \mathbf{0}, & \cdots & \mathbf{M}_{k} \mathbf{A}_{k} \end{array} \right)$$

$$\left[f\left(\left\langle \mathbf{A}_{v}^{\oplus}\right\rangle\right)\right]_{v}=\left[\left\langle \mathbf{M}_{i}\left\langle \mathbf{A}_{v}^{\oplus}\right\rangle_{i}|1\leq i\leq k\right\rangle\right]_{v}=\left(\begin{array}{c}\frac{\mathbf{M}_{1}\left\langle \mathbf{A}_{v}^{\oplus}\right\rangle_{1}}{\mathbf{M}_{2}\left\langle \mathbf{A}_{v}^{\oplus}\right\rangle_{2}}\\ \vdots\\ \overline{\mathbf{M}_{k}\left\langle \mathbf{A}_{v}^{\oplus}\right\rangle_{k}}\end{array}\right)=$$

$$= \left( \frac{\mathbf{M}_{1} \left( \mathbf{A}_{1}, \mathbf{0}, \cdots, \mathbf{0} \right)}{\mathbf{M}_{2} \left( \mathbf{0}, \mathbf{A}_{2}, \cdots, \mathbf{0} \right)} \right) = \left( \frac{\mathbf{M}_{1} \mathbf{A}_{1}, \mathbf{0}, \cdots, \mathbf{0}}{\mathbf{0}, \mathbf{M}_{2} \mathbf{A}_{2}, \cdots, \mathbf{0}} \right) \\ = \left( \frac{\mathbf{M}_{1} \mathbf{A}_{1}, \mathbf{0}, \cdots, \mathbf{0}}{\mathbf{0}, \mathbf{M}_{2} \mathbf{A}_{2}, \cdots, \mathbf{0}} \right) = f \left( \langle \mathbf{A} \rangle \right)_{v}^{\oplus}$$

For the case of right-multiplication we have:

$$f(\langle \mathbf{A} \rangle)_{h}^{\oplus} = \langle \mathbf{A}_{i} \mathbf{M}_{i} \rangle_{h}^{\oplus} = \begin{pmatrix} \mathbf{A}_{1} \mathbf{M}_{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{2} \mathbf{M}_{2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{k} \mathbf{M}_{k} \end{pmatrix}$$

$$[f(\langle \mathbf{A}_{h}^{\oplus} \rangle)]_{h} = [\langle \langle \mathbf{A}_{h}^{\oplus} \rangle_{i} \mathbf{M}_{i} | 1 \leq i \leq k \rangle]_{v} = [\langle \mathbf{A}_{v}^{\oplus} \rangle_{1} \mathbf{M}_{1} | \langle \mathbf{A}_{v}^{\oplus} \rangle_{2} \mathbf{M}_{2} | \cdots | \langle \mathbf{A}_{v}^{\oplus} \rangle_{k} \mathbf{M}_{k}] =$$

$$= \begin{pmatrix} \begin{bmatrix} \mathbf{A}_{1} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \mathbf{M}_{1} & \begin{bmatrix} \mathbf{0} \\ \mathbf{A}_{2} \\ \vdots \\ \mathbf{0} \end{bmatrix} \mathbf{M}_{2} & \cdots & \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \mathbf{M}_{k} & = \begin{pmatrix} \mathbf{A}_{1} \mathbf{M}_{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{2} \mathbf{M}_{2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \vdots & \ddots & \vdots \\ \mathbf{0} & \vdots & \ddots & \vdots \\ \mathbf{0} & \vdots & \ddots & \vdots \end{pmatrix} =$$

$$= f\left(\langle \mathbf{A} \rangle\right)_h^{\oplus}$$

B Blockwise Kronecker Product Factorisation for Vectors

**Proposition B1.** Let  $\mathbf{a}$  and  $\mathbf{b}$  be two column vectors, each comprised of k row blocks of arbitrary lengths. The blockwise Kronecker product  $\mathbf{a} \odot \mathbf{b}$  can be factorised as  $\mathbf{a} \odot \mathbf{b} = (\mathbf{a} \odot I_{\mathbf{b}}) \mathbf{b} = (I_{\mathbf{a}} \odot \mathbf{b}) \mathbf{a}$ , where  $I_{\mathbf{b}}$  and  $I_{\mathbf{a}}$  are identity matrices with the same number of rows and row block structure as  $\mathbf{b}$  and  $\mathbf{a}$ , respectively.

*Proof.* Let the vector **a** have dimensions  $M \times 1$  and  $\langle \mathbf{a} \rangle \equiv \left\langle \mathbf{a}_i \mid 1 \leq i \leq k \right\rangle$  be the sequence of vectors which represent the row blocks in **a** such that  $\sum_{i=1}^k m_i = M$ . Similarly, let the vector **b** have

dimensions  $N \times 1$  and the sequence of vectors  $\langle \boldsymbol{b} \rangle \equiv \left\langle \boldsymbol{b}_i \middle| 1 \le i \le k \right\rangle$  represent the row blocks in  $\boldsymbol{b}$  such that  $\sum_{i=1}^k n_i = N$ .

The blockwise Kronecker product  $\mathbf{a} \odot \mathbf{b}$  can then be defined as:

$$\mathbf{a}\odot oldsymbol{b} = egin{pmatrix} \mathbf{a}_1\otimes oldsymbol{b}_1\ \overline{\mathbf{a}_2\otimes oldsymbol{b}_2}\ \hline dots\ \overline{\mathbf{a}_k\otimes oldsymbol{b}_k} \end{pmatrix}$$

Note that by the properties of the Kronecker product the following holds for any two matrices  $\mathbf{A}$  and  $\mathbf{B}$ :

$$\mathbf{A}_{m \times p} \otimes \mathbf{B}_{n \times q} = (\mathbf{A} \otimes I_n) (I_p \otimes \mathbf{B}) = (I_m \otimes \mathbf{B}) (\mathbf{A} \otimes I_q)$$
(43)

In the special case of a Kronecker product between two vectors, p and q reduce to 1 and the identity matrices  $I_p$  and  $I_q$  become 1 by association. As a result, for any given Kronecker product  $\mathbf{a}_i \otimes \mathbf{b}_i$ , the following holds:

$$\mathbf{a}_{i} \otimes \mathbf{b}_{i} = (\mathbf{a}_{i} \otimes I_{p}) \, \mathbf{b}_{i} = (I_{m} \otimes \mathbf{b}_{i}) \, \mathbf{a}_{i} \tag{44}$$

This allows us to rewrite the blockwise Kronecker product  $\mathbf{a} \odot \mathbf{b}$  in two alternative ways:

$$\mathbf{a} \odot \mathbf{b} = \begin{pmatrix} \frac{(\mathbf{a}_1 \otimes I_{n_1}) \mathbf{b}_1}{(\mathbf{a}_2 \otimes I_{n_2}) \mathbf{b}_2} \\ \vdots \\ \hline (\mathbf{a}_k \otimes I_{n_k}) \mathbf{b}_k \end{pmatrix}$$
(45a)

$$\mathbf{a} \odot \mathbf{b} = \begin{pmatrix} \frac{(I_{m_1} \otimes \mathbf{b}_1) \mathbf{a}_1}{(I_{m_2} \otimes \mathbf{b}_2) \mathbf{a}_2} \\ \vdots \\ \hline (I_{m_k} \otimes \mathbf{b}_k) \mathbf{a}_k \end{pmatrix}$$
(45b)

Define a sequence of matrices  $\langle \mathbf{a}_i \otimes I_{n_i} | 1 \leq i \leq k \rangle$ , or  $\langle \mathbf{a}_i \otimes I_{n_i} \rangle$  for short, and another sequence  $\langle I_{m_i} \otimes \mathbf{b}_i | 1 \leq i \leq k \rangle \equiv \langle I_{m_i} \otimes \mathbf{b}_i \rangle$ . We can then rewrite (45a) and (45b) using the blockwise direct sum operator introduced in Section 2:

$$\mathbf{a} \odot \mathbf{b} = \langle \mathbf{a}_i \otimes I_{n_i} \rangle^{\oplus} \mathbf{b} \tag{46a}$$

$$\mathbf{a} \odot \mathbf{b} = \langle I_{m_i} \otimes \mathbf{b}_i \rangle^{\oplus} \mathbf{a} \tag{46b}$$

It then remains to show that  $\langle \mathbf{a}_i \otimes I_{n_i} \rangle^{\oplus}$  and  $\langle I_{m_i} \otimes \mathbf{b}_i \rangle^{\oplus}$  can be written as  $\mathbf{a} \odot I_{\mathbf{b}}$  and  $I_{\mathbf{a}} \odot \mathbf{b}$ , respectively. This can be done with the help of Proposition A1.

First of all, consider the sequences of identity matrices  $\langle I_{m_i} \rangle \equiv \langle I_{m_i} | 1 \leq i \leq k \rangle$  and  $\langle I_{n_i} \rangle \equiv \langle I_{n_i} | 1 \leq i \leq k \rangle$  on a standalone basis. The sequences  $\langle I_{m_i} \otimes \mathbf{b}_i \rangle$  and  $\langle \mathbf{a}_i \otimes I_{n_i} \rangle$  can then be expressed as functions  $f: S \to S$  and  $g: S \to S$  where S represents the set of all matrix sequences of length k, such that:

$$f\left(\langle I_{m_i}\rangle\right) = \left\langle\langle I_{m_i}\rangle_i \otimes \langle \boldsymbol{b}\rangle_i | 1 \leq i \leq k\right\rangle \equiv \langle I_{m_i} \otimes \boldsymbol{b}_i | 1 \leq i \leq k\right\rangle \equiv \langle I_{m_i} \otimes \boldsymbol{b}_i\rangle$$

$$g\left(\langle I_{n_i}\rangle\right) = \left\langle\langle \mathbf{a}\rangle_i \otimes \langle I_{n_i}\rangle_i | 1 \leq i \leq k\right\rangle \equiv \langle \mathbf{a}_i \otimes I_{n_i} | 1 \leq i \leq k\right\rangle \equiv \langle \mathbf{a}_i \otimes I_{n_i}\rangle$$

Proposition A1 applies because we can express  $\langle \mathbf{a}_i \otimes I_{n_i} \rangle$  and  $\langle I_{m_i} \otimes \mathbf{b}_i \rangle$  as left-multiplications over the sequences of identity matrcies  $\langle I_{m_i} \rangle$  and  $\langle I_{n_i} \rangle$ :

$$\langle I_{m_i} \otimes \boldsymbol{b}_i \rangle = \langle (I_{m_i} \otimes \boldsymbol{b}_i) I_{m_i} \rangle$$

$$\langle \mathbf{a}_i \otimes I_{n_i} \rangle = \langle (\mathbf{a}_i \otimes I_{n_i}) I_{n_i} \rangle$$

It then follows that:

$$f\left(\langle I_{m_i}\rangle\right)_v^{\oplus} = \left[f\left(\langle\langle I_{m_i}\rangle_v^{\oplus}\rangle\right)\right]_v$$

$$g\left(\langle I_{n_i}\rangle\right)_v^{\oplus} = \left[g\left(\langle\langle I_{n_i}\rangle_v^{\oplus}\rangle\right)\right]_v$$

Next, note that  $\langle I_{m_i} \rangle_v^{\oplus}$  produces an identity matrix of size M with a row block structure of  $\mathbf{a}$ , while  $\langle I_{n_i} \rangle_v^{\oplus}$  produces an identity matrix of size N with a row block structure of  $\mathbf{b}$ . In other words,  $\langle I_{m_i} \rangle_v^{\oplus} = I_{\mathbf{a}}$  and  $\langle I_{m_i} \rangle_v^{\oplus} = I_{\mathbf{b}}$ . This means:

$$\left[f\left(\left\langle \left\langle I_{m_i} \right\rangle_v^{\oplus} \right\rangle\right)\right]_v = \left[f\left(\left\langle I_{\mathbf{a}} \right\rangle\right)\right]_v = \left[\left\langle \left\langle I_{\mathbf{a}} \right\rangle_i \otimes \boldsymbol{b}_i \right\rangle\right]_v$$

$$\left[g\left(\left\langle \left\langle I_{n_{i}}\right\rangle _{v}^{\oplus}\right\rangle \right)\right]_{v}=\left[g\left(\left\langle I_{\boldsymbol{b}}\right\rangle \right)\right]_{v}=\left[\left\langle \mathbf{a}_{i}\otimes\left\langle I_{\boldsymbol{b}}\right\rangle _{i}\right\rangle \right]_{v}$$

Lastly, we note that the matrix representation of a pairwise Kronecker product over two sequences of matrices is the same as a blockwise Kronecker product if the sequences represent matrix blocks along the same dimension. For example,  $[\langle \mathbf{a}_i \otimes \langle I_b \rangle_i \rangle]_v = \mathbf{a} \odot I_b$  because  $\mathbf{a}_i$  and  $\langle I_b \rangle_i$  are row blocks of  $\mathbf{a}$  and  $I_b$ , respectively.

Putting all the steps together we have:

$$\mathbf{a} \odot \mathbf{b} = \langle \mathbf{a}_i \otimes I_{n_i} \rangle^{\oplus} \mathbf{b} = \left[ \left\langle \mathbf{a}_i \otimes \left\langle \left\langle I_{n_i} \right\rangle_v^{\oplus} \right\rangle_i \right\rangle \right]_v \mathbf{b} = \left[ \left\langle \mathbf{a}_i \otimes \left\langle I_{\mathbf{b}} \right\rangle_i \right\rangle \right]_v \mathbf{b} = \left( \mathbf{a} \odot I_{\mathbf{b}} \right) \mathbf{b}$$

$$(47)$$

$$\mathbf{a} \odot \mathbf{b} = \langle I_{m_i} \otimes \mathbf{b}_i \rangle^{\oplus} \mathbf{a} = \left[ \left\langle \left\langle \left\langle I_{m_i} \right\rangle_v^{\oplus} \right\rangle_i \otimes \mathbf{b}_i \right\rangle \right]_v \mathbf{a} = \left[ \left\langle \left\langle I_{\mathbf{b}} \right\rangle_i \otimes \mathbf{b} \right\rangle \right]_v \mathbf{a} = (I_{\mathbf{a}} \odot \mathbf{b}) \mathbf{a}$$
(48)

# C Derivation: Lagrange Multipliers for the Dependent

Recalling that  $\rho_y = (1 + \lambda_y)$  and  $\rho_l = \frac{\lambda_l}{2}$ , rewrite equation (26) as:

$$\rho_y \, \mathbf{\Sigma}_Y \mathbf{w} = \mathbf{v}_1 + \mathbf{v}_2 - \rho_l \mathbf{1}_n \tag{49}$$

To solve for  $\rho_l$ , pre-multiply both sides of (49) by  $\mathbf{1}'_n$  and re-arrange:

$$\rho_y \mathbf{1}'_n \mathbf{\Sigma}_Y \mathbf{w} = \mathbf{1}'_n (\mathbf{v}_1 + \mathbf{v}_2) - n\rho_l$$

$$n\rho_l = \mathbf{1}'_n (\mathbf{v}_1 + \mathbf{v}_2) - \rho_u \mathbf{1}'_n \mathbf{\Sigma}_Y \mathbf{w}$$

Dividing both sides by n produces:

$$\rho_l = \frac{1}{n} \mathbf{1}'_n \left( \mathbf{v}_1 + \mathbf{v}_2 \right) - \frac{1}{n} \rho_y \, \mathbf{1}'_n \mathbf{\Sigma}_Y \mathbf{w} \tag{50}$$

27

We can find an alternative solution for  $\rho_l$  by pre-multiplying both sides of (49) with  $\mathbf{w}'$  and recalling  $\mathbf{w}' \mathbf{\Sigma}_Y \mathbf{w} = \sigma_y^2$  and  $\mathbf{w}' \mathbf{1}_n = l_y$ :

$$\rho_y \sigma_y^2 = \mathbf{w}' \left( \mathbf{v}_1 + \mathbf{v}_2 \right) - \rho_l l_y$$

$$\rho_l l_y = \mathbf{w}' \left( \mathbf{v}_1 + \mathbf{v}_2 \right) - \rho_y \sigma_y^2$$

dividing both sides by  $l_y$  we get:

$$\rho_l = \frac{1}{l_y} \mathbf{w}' \left( \mathbf{v}_1 + \mathbf{v}_2 \right) - \frac{\sigma_y^2}{l_y} \rho_y \tag{51}$$

This alternative solution is not very practical because it does not allow for the case of  $l_y = 0$ . However, we can use it in conjunction with (50) to eliminate  $\rho_l$  and solve for  $\rho_y$ :

$$\frac{1}{n}\mathbf{1}'_{n}\left(\mathbf{v}_{1}+\mathbf{v}_{2}\right)-\frac{1}{n}\rho_{y}\mathbf{1}'_{n}\boldsymbol{\Sigma}_{Y}\mathbf{w}=\frac{1}{l_{y}}\mathbf{w}'\left(\mathbf{v}_{1}+\mathbf{v}_{2}\right)-\frac{\sigma_{y}^{2}}{l_{y}}\rho_{y}$$

$$\frac{\sigma_y^2}{l_y}\rho_y - \frac{1}{n}\rho_y \mathbf{1}_n' \mathbf{\Sigma}_Y \mathbf{w} = \frac{1}{l_y} \mathbf{w}' \left( \mathbf{v}_1 + \mathbf{v}_2 \right) - \frac{1}{n} \mathbf{1}_n' \left( \mathbf{v}_1 + \mathbf{v}_2 \right)$$

$$\rho_y \frac{n\sigma_y^2 - l_y \mathbf{1}_n' \mathbf{\Sigma}_Y \mathbf{w}}{nl_y} = \frac{(n\mathbf{w} - l_y \mathbf{1}_n)' (\mathbf{v}_1 + \mathbf{v}_2)}{nl_y}$$

Rearranging for  $\rho_y$  we get:

$$\rho_y = \frac{(n\mathbf{w} - l_y \mathbf{1}_n)'(\mathbf{v}_1 + \mathbf{v}_2)}{n\sigma_y^2 - l_y \mathbf{1}_n' \mathbf{\Sigma}_Y \mathbf{w}}$$
(52)

# D Derivation: Lagrange Multipliers for the Explanatory

Start from the first-order condition for  $\omega$ . Expressing in terms of  $\lambda_p$ , we get:

$$\mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \boldsymbol{\lambda}_{p} = 2\mathbf{v}_{3} - 2\mathbf{v}_{4} - 2\left(\mathbf{u} \odot I_{\boldsymbol{\omega}}\right)' \boldsymbol{\Sigma}_{X}^{d} \left(\mathbf{u} \odot I_{\boldsymbol{\omega}}\right) \boldsymbol{\omega}^{\oplus} \boldsymbol{\lambda}_{x}$$

$$(53)$$

Define the following shorthand notations for convenience:

$$\begin{split} \mathbf{\Theta}_{K\times K} &= \operatorname{diag}(\boldsymbol{\vartheta}_x) = \begin{pmatrix} \sigma_{x,1}^2, & 0, & \cdots & 0 \\ 0, & \sigma_{x,2}^2, & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \cdots & \sigma_{x,K}^2 \end{pmatrix}, \\ \mathbf{L}_{K\times K} &= \operatorname{diag}(\boldsymbol{l}_p) = \begin{pmatrix} l_{p,1}, & 0, & \cdots & 0 \\ 0, & l_{p,2}, & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \cdots & l_{p,K} \end{pmatrix}, \end{split}$$

$$\mathbf{M}_{1} = (\mathbf{1}_{\boldsymbol{\omega}}^{\oplus})' \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} = \begin{pmatrix} m_{1}, & 0, & \cdots & 0 \\ 0, & m_{2}, & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \cdots & m_{K} \end{pmatrix}, \quad \mathbf{M}_{2} = (\mathbf{u} \odot I_{\boldsymbol{\omega}})' \boldsymbol{\Sigma}_{X}^{d} (\mathbf{u} \odot I_{\boldsymbol{\omega}})$$

A system of two vector equations, each with K rows and K unknowns, is produced by pre-multiplying the first-order condition for  $\boldsymbol{\omega}$  with  $(\boldsymbol{\omega}^{\oplus})'$  and  $(\mathbf{1}_{\boldsymbol{\omega}}^{\oplus})'$ , respectively:

$$\begin{cases}
\mathbf{L}\boldsymbol{\lambda}_{p} = 2\left(\boldsymbol{\omega}^{\oplus}\right)'(\mathbf{v}_{3} - \mathbf{v}_{4}) - 2\boldsymbol{\Theta}\boldsymbol{\lambda}_{x} \\
\mathbf{M}_{1}\boldsymbol{\lambda}_{p} = 2\left(\mathbf{1}_{\boldsymbol{\omega}}^{\oplus}\right)'(\mathbf{v}_{3} - \mathbf{v}_{4}) - 2\left(\mathbf{1}_{\boldsymbol{\omega}}^{\oplus}\right)'\mathbf{M}_{2}\boldsymbol{\omega}^{\oplus}\boldsymbol{\lambda}_{x}
\end{cases} (54a)$$

Pre-multiplying 54a and 54b by  $\mathbf{L}^{-1}$  and  $\mathbf{M}_1^{-1}$ , respectively, we get:

$$\begin{cases}
\boldsymbol{\lambda}_{p} = 2\mathbf{L}^{-1} \left(\boldsymbol{\omega}^{\oplus}\right)' (\mathbf{v}_{3} - \mathbf{v}_{4}) - 2\mathbf{L}^{-1}\boldsymbol{\Theta}\boldsymbol{\lambda}_{x} \\
\boldsymbol{\lambda}_{p} = 2\mathbf{M}_{1}^{-1} \left(\mathbf{1}_{\boldsymbol{\omega}}^{\oplus}\right)' (\mathbf{v}_{3} - \mathbf{v}_{4}) - 2\mathbf{M}_{1}^{-1} \left(\mathbf{1}_{\boldsymbol{\omega}}^{\oplus}\right)' \mathbf{M}_{2}\boldsymbol{\omega}^{\oplus}\boldsymbol{\lambda}_{x}
\end{cases} (55a)$$

The solution for  $\lambda_p$  given by 55b is more practical because 55a does not allow for the case of zero-sum weights (zeros on the diagonal of **L** would make it uninvertible).

The solution for  $\lambda_x$  can be derived by setting the right-hand side of 55a equal to the right-hand side of 55b:

$$2\mathbf{L}^{-1}\left(\boldsymbol{\omega}^{\oplus}\right)'(\mathbf{v}_{3}-\mathbf{v}_{4})-2\mathbf{L}^{-1}\boldsymbol{\Theta}\boldsymbol{\lambda}_{x}=2\mathbf{M}_{1}^{-1}\left(\mathbf{1}_{\boldsymbol{\omega}}^{\oplus}\right)'(\mathbf{v}_{3}-\mathbf{v}_{4})-2\mathbf{M}_{1}^{-1}\left(\mathbf{1}_{\boldsymbol{\omega}}^{\oplus}\right)'\mathbf{M}_{2}\boldsymbol{\omega}^{\oplus}\boldsymbol{\lambda}_{x}$$

We can pre-multiply both sides by  $\mathbf{M}_1\mathbf{L}$  to avoid problems with inverting  $\mathbf{L}$  in the presence of zero-sum weight constraints (note that  $\mathbf{M}_1\mathbf{L} = \mathbf{L}\mathbf{M}_1$  because both are diagonal):

$$2\mathbf{M}_{1}\left(\boldsymbol{\omega}^{\oplus}\right)'(\mathbf{v}_{3}-\mathbf{v}_{4})-2\mathbf{M}_{1}\boldsymbol{\Theta}\boldsymbol{\lambda}_{x}=2\mathbf{L}\left(\mathbf{1}_{\boldsymbol{\omega}}^{\oplus}\right)'(\mathbf{v}_{3}-\mathbf{v}_{4})-2\mathbf{L}\left(\mathbf{1}_{\boldsymbol{\omega}}^{\oplus}\right)'\mathbf{M}_{2}\boldsymbol{\omega}^{\oplus}\boldsymbol{\lambda}_{x}$$

Re-arranging for  $\lambda_x$  we get:

$$\lambda_{x} = \left[ \mathbf{M}_{1} \mathbf{\Theta} - \mathbf{L} \left( \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \right)' \mathbf{M}_{2} \boldsymbol{\omega}^{\oplus} \right]^{-1} \left( \boldsymbol{\omega}^{\oplus} \mathbf{M}_{1} - \mathbf{1}_{\boldsymbol{\omega}}^{\oplus} \mathbf{L} \right)' (\mathbf{v}_{3} - \mathbf{v}_{4})$$
 (56)

### E Data reference

Table 1: Data series used in the empirical study

Dataset	Source	$\mathrm{Ticker}^1$	Frequency	History start
Real GDP	U.S. Bureau of Economic Analysis	GDPC1	Quarterly	1947Q1
Personal Consumption Expenditure (Cons.)	U.S. Bureau of Economic Analysis	PCECC96	Quarterly	1947Q1
Gross Private Domestic Investment (Inv.)	U.S. Bureau of Economic Analysis	GPDIC1	Quarterly	1947Q1
Government Consumption and Investment (Govt.)	U.S. Bureau of Economic Analysis	GCEC1	Quarterly	1947Q1
Exports of Goods and Services (Exp.)	U.S. Bureau of Economic Analysis	EXPGSC1	Quarterly	1947Q1
Imports of Goods and Services (Imp.)	U.S. Bureau of Economic Analysis	IMPGSC1	Quarterly	1947Q1
S&P 500	Investing.com	US500	Monthly	1989-10
Energy (En.)	Investing.com	SPNY	Monthly	1989-10
Materials (Mat.)	Investing.com	SPLRCM	Monthly	1989-10
Industrials (Ind.)	Investing.com	SPLRCI	Monthly	1989-10
Financials (Fin.)	Investing.com	SPSY	Monthly	1989-10
Healthcare (HC)	Investing.com	SPXHC	Monthly	1989-10
Consumer Discretionary (Disc.)	Investing.com	SPLRCD	Monthly	1989-10
Consumer Staples (Stapl.)	Investing.com	SPLRCS	Monthly	1989-10
Communication (Telco)	Investing.com	SPLRCL	Monthly	1989-10
Technology (IT)	Investing.com	SPLRCT	Monthly	1989-10
Utilities (Util.)	Investing.com	SPLRCU	Monthly	1989-10

<sup>&</sup>lt;sup>1</sup> Data for U.S. GDP and its individual expenditure components was retrieved from the St. Louis Federal Reserve economic database (FRED) on 25 April 2025. The corresponding tickers are identifiers for the FRED database. Data for the S&P 500 and its sector sub-indices was retrieved directly from Investing.com on 25 April 2025 using the tickers above.