# Dynamic Mode Decompositions and Vector Autoregressions *

Thomas J. Sargent, Yatheesan J. Selvakumar, and Ziyue Yang

January 8, 2026

**Abstract**

We establish connections between Dynamic Mode Decompositions (DMDs), vector autoregressions, and linear state-space models, showing that DMD efficiently estimates low-rank first-order VAR projection coefficients from high-dimensional data. When the measurement matrix has full column rank, the recovered nonzero eigenvalues coincide with those of the underlying state transition matrix. We apply DMD to a 100-household heterogeneous-agent economy with complete markets and Gorman aggregation. From high-dimensional household income and consumption panels, DMD successfully recovers low-dimensional aggregate dynamics: estimated modes track latent aggregate states with correlations exceeding 0.90, and cross-sectional loadings reveal the sharing rule governing redistribution. This demonstrates DMD's capacity to extract economically meaningful low-dimensional structure from microeconomic panels.

**Keywords:** Vector Autoregression, LQG State-Space Model, Innovations Representation, Singular Value Decomposition, Dynamic Mode Decomposition, Information Projection

*For critical comments and suggestions, we thank participants at the May 2024 Atlanta Fed conference in honor of Christopher A. Sims, especially Marco Bassetto and Mikkel Plagborg-Moller.

# 1 Introduction

Applied mathematicians have recently used Dynamic Mode Decompositions (DMDs) as a "machine learning" technique to quantify non-random dynamical systems, including fluid dynamics (e.g., see Brunton and Kutz (2022)). This paper instead uses DMDs to quantify a dynamical system that is persistently buffeted by random shocks. We describe connections among Dynamic Mode Decompositions (DMDs), vector autoregressions (VARs), and a special case of what Stock and Watson (2016) call dynamic factor models. DMDs are calculated using singular value decompositions and associated eigen decompositions, but of different objects than the ones that underlie the dynamic factor models described by Stock and Watson. In this paper we show that, for an observation-noise-free linear state-space model, the population first-order VAR projection coefficient is reduced-rank and can be efficiently estimated and represented by DMD. We use a Dynamic Mode Decomposition (DMD) to infer a reduced-rank first-order VAR from a data set $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T, \mathbf{y}_{T+1}]$.

Through its connection to the Koopman operator, the DMD algorithm has also been used to approximate nonlinear dynamics (see Williams et al. (2015), Brunton et al. (2016), Mezic (2020)). Our model resembles those of Geweke (1977), Sargent and Sims (1977), Geweke and Singleton (1981), Stock and Watson (2002), Bai (2003), and Bai and Ng (2013), but our statistical model of hidden factors differs from theirs. Like those authors, we estimate "non-structural Kepler-stage" descriptive models that can compress data and reveal patterns. As Koopmans (1947) recommends, we want eventually to interpret these empirical regularities with a "structural Newton-stage" model that is cast in terms of parameters that describe market structures and decision makers' preferences, constraints, and information flows. Koopmans interpreted Burns and Mitchell (1946) as such a "Kepler" stage model of business cycles, in contrast to the structural, simultaneous stochastic difference equation models of business cycles that

could be constructed with tools developed by Koopmans (1950), Hood and Koopmans (1953), and Marschak (1953).

This paper is organized as follows. Section 2 describes our statistical model, its relationship to a reduced-rank first-order VAR, and its connections to Dynamic Mode Decompositions. This section focuses on five objects:

1. A statistical model for a covariance-stationary stochastic process $\{\mathbf{y}_t\}_{t=-\infty}^{\infty}$ that takes the form of a special observation-noise-free LQG state-space model, where $\mathbf{y}_t$ is an $M \times 1$ vector.
2. An innovations representation of that state-space model.
3. An infinite-order vector autoregression (VAR) for $\{\mathbf{y}_t\}_{t=-\infty}^{\infty}$.
4. A reduced-rank first-order VAR.
5. A Dynamic Mode Decomposition (DMD) of a data set $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T, \mathbf{y}_{T+1}]$.

Section 2 describes situations in which

- items [1] and [2] are both valid representations of item [3]
- items [3] and [4] coincide
- item [5] provides a good way to estimate item [4]

Having acknowledged that we understand that a DMD provides a descriptive statistical model, not what theoretical IO or macroeconomists call a "structural" model, Section 3 proceeds to analyze the qualities of our Section 2 model as a statistical approximation to a particular structural dynamic stochastic general equilibrium (DSGE) model. The model describes competitive equilibrium prices and quantities for a heterogeneous-agent economy with what Milton Friedman called "homogenizing mechanisms" that can put a factor structure into the dynamics of cross-sections of households' income and consumption rates. We design the DSGE model so that its equilibrium has a representation that takes the form of a linear state-space model and an associated likelihood function that can readily be compared via "information projection" techniques to its

DMD counterpart for the same observable process.[1] The computational experiments in our Section 3 laboratory are designed to set the stage for an empirical application to US CEX data in Sargent et al. (2025). Section 4 describes differences between DMDs and some other dynamic factor models that are widely used by economists. Section 5 offers concluding remarks about promising applications.

## 2 Statistical Model Underlying DMD

Let $\mathbf{y}_t$ be an $M \times 1$ vector and let $\mathbf{y}^{t-1} := \{\mathbf{y}_{t-s}\}_{s=1}^{\infty}$ denote the full past. Throughout we use $\perp$ to denote orthogonality and work with centered observables with $\mathbb{E}[\mathbf{y}_t] = 0$.

We take as baseline an observation-noise-free linear state-space model with potentially rank-deficient measurement:

$$\begin{aligned} \mathbf{x}_{t+1} &= \underset{N_x \times N_x}{\mathbf{A}}\, \mathbf{x}_t + \underset{N_x \times N_x}{\mathbf{C}}\, \mathbf{w}_{t+1} \\ \mathbf{y}_t &= \underset{M \times N_x}{\mathbf{G}}\, \mathbf{x}_t, \end{aligned} \tag{1}$$

where random shocks $\mathbf{w}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_x})$. Let $r := \text{rank}(\mathbf{G}) \leq \min\{M, N_x\}$. When $r < M$, $\{\mathbf{y}_t\}$ is supported on the $r$-dimensional subspace $\text{range}(\mathbf{G}) \subset \mathbb{R}^M$ and $\text{Cov}(\mathbf{y}_t)$ is singular. Equivalently, writing (1) at time $t$ gives $\mathbf{x}_t = \mathbf{A}\,\mathbf{x}_{t-1} + \mathbf{C}\,\mathbf{w}_t$, so the innovation $\mathbf{w}_t$ drives $\mathbf{y}_t$. Assume $\mathbf{A}$ is stable so that $\{\mathbf{x}_t\}$ and $\{\mathbf{y}_t\}$ are covariance-stationary. The optimal one-step-ahead prediction conditioned on $\mathbf{x}_t$ is

$$\check{\mathbf{x}}_{t+1} = \mathbb{E}[\mathbf{x}_{t+1} \mid \mathbf{x}_t] = \mathbf{A}\,\mathbf{x}_t,$$

and the associated one-step-ahead prediction error covariance matrix is

$$\mathbb{E}\left[(\mathbf{x}_{t+1} - \check{\mathbf{x}}_{t+1})(\mathbf{x}_{t+1} - \check{\mathbf{x}}_{t+1})^\top\right] = \mathbf{C}\,\mathbf{C}^\top.$$

[1] See footnote 13 below.

The associated steady-state innovations representation is cast in terms of the one-step-ahead prediction vector

$$\widehat{\mathbf{x}}_{t+1} = \mathbb{E}[\mathbf{x}_{t+1} \,|\, \mathbf{y}^t].$$

This innovations representation is

$$\begin{aligned} \underset{N_x \times 1}{\widehat{\mathbf{x}}_{t+1}} &= \underset{N_x \times N_x}{\mathbf{A}}\, \widehat{\mathbf{x}}_t + \underset{N_x \times M}{\mathbf{K}}\, \mathbf{a}_t, \\ \underset{M \times 1}{\mathbf{y}_t} &= \underset{M \times N_x}{\mathbf{G}}\, \widehat{\mathbf{x}}_t + \mathbf{a}_t, \end{aligned} \tag{2}$$

where $\mathbf{a}_t = \mathbf{y}_t - \mathbb{E}[\mathbf{y}_t \,|\, \mathbf{y}^{t-1}]$, $\mathbf{a}_t \perp \mathbf{y}^{t-1}$, and the steady-state Kalman gain is

$$\mathbf{K} = \mathbf{A}\, \mathbf{\Sigma}_\infty\, \mathbf{G}^\top (\mathbf{G}\, \mathbf{\Sigma}_\infty\, \mathbf{G}^\top)^+, \tag{3}$$

where $(\cdot)^+$ denotes the Moore–Penrose pseudoinverse[2], and

$$\mathbf{\Sigma}_\infty = \mathbb{E}\left[(\mathbf{x}_t - \widehat{\mathbf{x}}_t)(\mathbf{x}_t - \widehat{\mathbf{x}}_t)^\top\right]$$

is the positive semidefinite solution of the following algebraic matrix Riccati equation:

$$\mathbf{\Sigma}_\infty = \mathbf{C}\, \mathbf{C}^\top + (\mathbf{A} - \mathbf{K}\, \mathbf{G}) \mathbf{\Sigma}_\infty (\mathbf{A} - \mathbf{K}\, \mathbf{G})^\top. \tag{4}$$

The population lag-1 linear projection coefficient is

$$\mathbf{B}^* = \mathrm{Cov}(\mathbf{y}_t, \mathbf{y}_{t-1})\ \mathrm{Cov}(\mathbf{y}_{t-1})^+. \tag{5}$$

Note that when $\mathrm{Cov}(\mathbf{y}_{t-1})$ is singular, there are many matrices that yield the same fitted values on $\mathrm{range}(\mathrm{Cov}(\mathbf{y}_{t-1}))$. We adopt the unique Moore–Penrose representative.

[2]We use a pseudoinverse in (3) because $\mathbf{G}\, \mathbf{\Sigma}_\infty\, \mathbf{G}^\top$ has rank at most $r := \mathrm{rank}(\mathbf{G})$ and is singular when $r < M$.

Define the associated linear prediction error $\boldsymbol{\varepsilon}_t := \mathbf{y}_t - \mathbf{B}^* \mathbf{y}_{t-1}$, so that

$$\mathbf{y}_t = \underset{M\times M}{\mathbf{B}^*} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{6}$$
$$\mathbb{E}[\boldsymbol{\varepsilon}_t \mathbf{y}_{t-1}^\top] = \mathbf{0}.$$

When the restriction $\mathbb{E}[\boldsymbol{\varepsilon}_t \mathbf{y}_{t-j}^\top] = \mathbf{0}$ holds for all $j \geq 1$, (6) becomes a VAR(1) with innovations orthogonal to the full past. Let $\boldsymbol{\Sigma}_x := \mathrm{Cov}(\mathbf{x}_t)$ denote the stationary state covariance that solves the Lyapunov equation

$$\boldsymbol{\Sigma}_x = \mathbf{A}\,\boldsymbol{\Sigma}_x\,\mathbf{A}^\top + \mathbf{C}\,\mathbf{C}^\top.$$

The next proposition characterizes the lag-1 linear projection coefficient $\mathbf{B}^*$ and its relationship to the state transition matrix $\mathbf{A}$.

**Proposition 1.** *Let $\boldsymbol{\Sigma}_x := \mathrm{Cov}(\mathbf{x}_t)$ under (1) and let $\mathbf{B}^*$ be defined by (5). Then*

*(i) $\mathbf{B}^*$ satisfies*

$$\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\boldsymbol{\Sigma}_x\,\mathbf{G}^\top\big(\mathbf{G}\,\boldsymbol{\Sigma}_x\,\mathbf{G}^\top\big)^+ = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^+_{\boldsymbol{\Sigma}_x}, \quad \mathbf{G}^+_{\boldsymbol{\Sigma}_x} := \boldsymbol{\Sigma}_x\,\mathbf{G}^\top\big(\mathbf{G}\,\boldsymbol{\Sigma}_x\,\mathbf{G}^\top\big)^+.$$

*(ii) The following properties hold:*

$$\mathrm{range}(\mathbf{B}^*) \subseteq \mathrm{range}(\mathbf{G}), \quad \mathit{rank}(\mathbf{B}^*) \leq \mathit{rank}(\mathbf{G}),$$
$$\mathbf{B}^*\,\mathbf{v} = \lambda\,\mathbf{v},\ \lambda \neq 0 \Rightarrow \mathbf{v} \in \mathrm{range}(\mathbf{G}).$$

*Proof.* See Appendix A.1.1. □

Proposition 1 implies that, without further structure, the object identified from $\{\mathbf{y}_t\}$ is $\mathbf{B}^*$, the best linear predictor of $\mathbf{y}_t$ given $\mathbf{y}_{t-1}$. What this reveals about the underlying state dynamics depends on whether the measurement map $\mathbf{x}_t \mapsto \mathbf{y}_t = \mathbf{G}\,\mathbf{x}_t$ is injective.

If $\text{rank}(\mathbf{G}) < N_x$, then $\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^+_{\mathbf{\Sigma}_x}$ depends on $\mathbf{A}$ only through its transformation on the $\mathbf{G}$-visible subspace, $\mathbf{A}$ is not identified from $\mathbf{B}^*$ without additional restrictions, and $\mathbf{B}^*$ represents only the linear projection coefficient, which may differ from a VAR(1) coefficient if the innovations are not white. This shows that full column rank of $\mathbf{G}$ is a key condition for recovering $\mathbf{A}$ from $\mathbf{B}^*$, and the following proposition further explores implications of this condition.

**Proposition 2.** *Suppose that* $\mathbf{G}$ *has full column rank and let* $\mathbf{G}^+ = (\mathbf{G}^\top \mathbf{G})^{-1}\,\mathbf{G}^\top$. *Define*

$$\widetilde{\mathbf{B}} := \mathbf{G}\,\mathbf{A}\,\mathbf{G}^+, \qquad \mathbf{u}_t := \mathbf{G}\,\mathbf{C}\,\mathbf{w}_t\,.$$

*Then the following statements hold:*

*(i)* $\{\mathbf{y}_t\}$ *admits VAR(1) representation*

$$\mathbf{y}_t = \widetilde{\mathbf{B}}\,\mathbf{y}_{t-1} + \mathbf{u}_t,$$

*where* $\{\mathbf{u}_t\}$ *is i.i.d. with* $\mathbb{E}[\mathbf{u}_t] = \mathbf{0}$ *and* $\mathbf{u}_t$ *is independent of the past* $\mathbf{y}^{t-1}$. *Moreover,* $\mathbf{a}_t$ *in* (2) *satisfies* $\mathbf{a}_t = \mathbf{u}_t$, *and* $\mathbf{B}^*$ *in* (5) *satisfies*

$$\mathbf{B}^*\,\mathbf{y}_{t-1} = \widetilde{\mathbf{B}}\,\mathbf{y}_{t-1} \qquad a.s.$$

*(ii) If in addition* $\mathbf{\Sigma}_x := \text{Cov}(\mathbf{x}_t) \succ 0$, *then*

$$\mathbf{B}^* = \widetilde{\mathbf{B}} = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^+, \qquad \mathbf{A} = \mathbf{G}^+\,\mathbf{B}^*\,\mathbf{G}\,.$$

*Proof.* See Appendix A.1.2. □

Proposition 2 highlights what can be inferred about $\mathbf{A}$ from the lag-1 linear projection coefficient once the measurement map is injective. When $\mathbf{G}$ has full column

rank and $\mathbf{\Sigma}_x \succ 0$ (so that $\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^+$ by Proposition 2(ii)), the same algebra with any symmetric positive definite $N_x \times N_x$ matrix $\mathbf{S}$ in place of $\mathbf{\Sigma}_x$ shows that we can equivalently write

$$\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\mathbf{S}\,\mathbf{G}^\top(\mathbf{G}\,\mathbf{S}\,\mathbf{G}^\top)^+. \tag{7}$$

Moreover, Corollary 1 below shows that under $\mathbf{C}\,\mathbf{C}^\top \succ 0$, the Kalman gain and state transition matrix satisfy $\mathbf{A} = \mathbf{K}\,\mathbf{G}$.

**Corollary 1.** *Suppose that* $\mathbf{G}$ *has full column rank. If in addition* $\mathbf{C}\,\mathbf{C}^\top \succ 0$*, then the Riccati equation* (4) *reduces to*

$$\mathbf{\Sigma}_\infty = \mathbf{C}\,\mathbf{C}^\top \succ 0, \tag{8}$$

*and the corresponding Kalman gain satisfies* $\mathbf{A} = \mathbf{K}\,\mathbf{G}$.

*Proof.* See Appendix A.1.3. □

If in addition $\mathbf{C}\,\mathbf{C}^\top \succ 0$, Corollary 1 implies that $\mathbf{\Sigma}_\infty = \mathbf{C}\,\mathbf{C}^\top \succ 0$, so we can choose $\mathbf{S} = \mathbf{\Sigma}_\infty$ and write (7) as

$$\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top\big(\mathbf{G}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top\big)^+. \tag{9}$$

Comparing (9) with the Kalman-gain formula (3), we obtain

$$\mathbf{B}^* = \mathbf{G}\,\mathbf{K}\,.$$

To summarize:

(i) Under full column rank $\mathbf{G}$, Proposition 2(i) gives a VAR(1) with innovations $\mathbf{u}_t = \mathbf{G}\,\mathbf{C}\,\mathbf{w}_t$ orthogonal to the full past, and the first-order coefficient is uniquely represented by the Moore–Penrose choice $\mathbf{B}^*$ in (5).

(ii) Under full column rank $\mathbf{G}$ and $\mathbf{\Sigma}_x \succ 0$, Proposition 2(ii) yields $\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^+$ and $\mathbf{A} = \mathbf{G}^+\,\mathbf{B}^*\,\mathbf{G}$. In particular, the nonzero eigenvalues of $\mathbf{B}^*$ equal the eigenvalues

of $\mathbf{A}$ (with additional zeros when $M > N_x$).

(iii) Under full column rank $\mathbf{G}$ and $\mathbf{C}\,\mathbf{C}^\top \succ 0$, Corollary 1 yields $\mathbf{A} = \mathbf{K}\,\mathbf{G}$ and $\mathbf{\Sigma}_\infty = \mathbf{C}\,\mathbf{C}^\top$, and $\mathbf{B}^* = \mathbf{G}\,\mathbf{K}$.

### 2.1 DMD as Estimator of Reduced-Rank First-Order VAR

In this section, we show how Dynamic Mode Decomposition (DMD) provides a computationally efficient rank-$N$ estimator $\widehat{\mathbf{B}}$ of $\mathbf{B}^*$ from data matrices. When $\mathbf{G}$ has full column rank and $\mathbf{\Sigma}_x \succ 0$, Proposition 2(ii) implies that the nonzero eigenvalues of $\mathbf{B}^*$ equal those of $\mathbf{A}$ and $\mathbf{A} = \mathbf{G}^+\,\mathbf{B}^*\,\mathbf{G}$ if $\mathbf{G}$ is known, so replacing $\mathbf{B}^*$ by $\widehat{\mathbf{B}}$ yields an estimator of $\mathbf{A}$. If $\mathbf{G}$ is unknown, DMD estimates a low-dimensional invariant subspace of $\mathbf{B}^*$ encoded by its dynamic modes.

We estimate the lag-1 projection coefficient $\mathbf{B}^*$ in (6) using a data set organized as follows. Let $\mathbf{y}_t$ denote an $M \times 1$ vector of demeaned random variables for $t = 1, \ldots, T+1$,[3] and assume $M > T$, so there are more variables than time periods. Stack the $\mathbf{y}_t$ across time to create two $M \times T$ data matrices $\mathbf{Y}$ and $\mathbf{Y}'$:

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T], \tag{10}$$

$$\mathbf{Y}' = [\mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_{T+1}]. \tag{11}$$

Thus, we have $MT$ data points from which we want to estimate $M^2$ parameters in an $M \times M$ coefficient matrix. When $M^2 > MT$, least squares is underdetermined. Hence, we impose a rank constraint on the estimator. Fix a target reduced rank $N \leq \min\{M, T\}$. We target the rank-constrained least-squares problem

$$\widehat{\mathbf{B}} = \underset{\text{rank}(\mathbf{B}) \leq N}{\arg\min} \; \| \mathbf{Y}' - \mathbf{B}\,\mathbf{Y} \|_F^2. \tag{12}$$

[3] Hirsh et al. (2020) show that centering the data is equivalent to incorporating an affine term in the dynamic model and improves the performance of DMD in correctly extracting the dynamics of the data.

Assume the smallest singular value $\sigma_N > 0$ so the reduced SVD below exists. Compute the reduced SVD of $\mathbf{Y}$ with $\mathbf{U} \in \mathbb{R}^{M\times N}$, $\mathbf{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_N) \in \mathbb{R}^{N\times N}$, and $\mathbf{V} \in \mathbb{R}^{T\times N}$:

$$\mathbf{Y} \approx \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^\top, \tag{13}$$

which is the best rank-$N$ approximation of $\mathbf{Y}$ in the Frobenius norm by the Eckart–Young–Mirsky theorem (Golub and Van Loan, 2013, Theorem 2.4.8). Define the rank-$N$ approximation $\mathbf{Y}_N := \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^\top$. The solution to the rank-constrained problem (12) can be written using the Moore–Penrose pseudoinverse of $\mathbf{Y}_N$.

With (13), the Moore–Penrose pseudoinverse of the rank-$N$ truncation $\mathbf{Y}_N$ is

$$\mathbf{Y}_N^+ = \mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{U}^\top, \quad \text{so that} \quad \mathbf{Y}_N^+\,\mathbf{Y}_N = \mathbf{V}\,\mathbf{V}^\top,$$

since $\mathbf{Y}_N = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^\top$ and $\mathbf{U}^\top\,\mathbf{U} = \mathbf{I}_N$ imply $\mathbf{Y}_N^+\,\mathbf{Y}_N = \mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{U}^\top\,\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^\top = \mathbf{V}\,\mathbf{V}^\top$. We use this to form the least-squares estimator on the rank-$N$ subspace:

$$\widehat{\mathbf{B}} := \mathbf{Y}'\,\mathbf{Y}_N^+ = \mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{U}^\top.$$

Define the reduced operator

$$\widetilde{\mathbf{A}} := \mathbf{U}^\top\,\mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1},$$

with eigendecomposition $\widetilde{\mathbf{A}}\,\mathbf{W} = \mathbf{W}\,\mathbf{\Lambda}$, and set

$$\mathbf{\Phi} := \mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{W}.$$

The columns of $\mathbf{\Phi}$ are the DMD dynamic modes. The following proposition summarizes key properties of $\widehat{\mathbf{B}}$ and $\mathbf{\Phi}$.

**Proposition 3.** *Let* $\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}$ *be as in* (13) *and let* $\mathbf{P}_{\mathbf{\Phi}} := \mathbf{\Phi}\mathbf{\Phi}^+$ *denote the orthogonal projection onto the column space of* $\mathbf{\Phi}$*. Then the following hold:*

*(i)* $\widehat{\mathbf{B}}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda}$

*(ii)* $\widehat{\mathbf{B}}\mathbf{P}_{\mathbf{\Phi}} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^{+}$.

*Proof.* See Appendix A.1.4. □

Given $\widehat{\mathbf{B}}$, define residuals $\widehat{\mathbf{a}}_t := \mathbf{y}_t - \widehat{\mathbf{B}}\,\mathbf{y}_{t-1}$ for $t = 2, \ldots, T+1$ and the sample covariance $\widehat{\mathbf{\Omega}} = T^{-1}\sum_{t=2}^{T+1}\widehat{\mathbf{a}}_t\widehat{\mathbf{a}}_t^{\top}$. These residuals satisfy the sample first-order normal equations (Proposition 4) and are therefore pseudo-innovations for the fitted VAR(1), not necessarily innovations for the data-generating process. Proposition 3 states that on the rank-$N$ fitted subspace, $\widehat{\mathbf{B}}\mathbf{P}_{\mathbf{\Phi}} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^{+}$.

Proposition 1 implies that the population lag-1 projection coefficient satisfies $\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^{+}_{\mathbf{\Sigma}_x}$, with $\mathrm{range}(\mathbf{B}^*) \subseteq \mathrm{range}(\mathbf{G})$ and all eigenvectors of $\mathbf{B}^*$ associated with nonzero eigenvalues lying in $\mathrm{range}(\mathbf{G})$. Proposition 3 shows that DMD produces a rank-$N$ estimator $\widehat{\mathbf{B}}$ whose spectral information is encoded in its dynamic modes $\mathbf{\Phi}$ and eigenvalues $\mathbf{\Lambda}$. Under standard eigen-gap conditions, the column space of $\mathbf{\Phi}$ estimates an invariant subspace of $\mathbf{B}^*$ and the corresponding eigenvalues in $\mathbf{\Lambda}$ estimate the associated nonzero eigenvalues of $\mathbf{B}^*$.[4] When $\mathbf{G}$ has full column rank and $\mathbf{\Sigma}_x \succ 0$, these nonzero eigenvalues equal the eigenvalues of the underlying state transition matrix $\mathbf{A}$ (Proposition 2(ii)). When $\mathbf{G}$ is rank-deficient, $\mathbf{\Lambda}$ should be interpreted as describing the eigenvalues of the observed lag-1 projection operator $\mathbf{B}^*$. These eigenvalues coincide with a subset of eigenvalues of $\mathbf{A}$ only under additional structure (e.g., when the $\mathbf{G}$-visible state subspace is invariant under $\mathbf{A}$).

Algorithm 1 summarizes the exact DMD procedure for estimating the reduced-rank VAR operator $\widehat{\mathbf{B}}$, DMD modes $\mathbf{\Phi}$, and eigenvalues $\mathbf{\Lambda}$ from an $M \times (T+1)$ data matrix $\mathbf{Y}$.

[4] Eigenvector convergence typically requires well-separated eigenvalues. When two eigenvalues are close or repeated, the corresponding eigenspaces converge but individual eigenvectors within those spaces may not be uniquely identified.

**Algorithm 1** Exact DMD for reduced-rank VAR estimation (Tu et al., 2014)

**Input:** Data matrix $\mathbf{Y} \in \mathbb{R}^{M \times (T+1)}$, rank $N$

1. Demean rows: $\widetilde{\mathbf{Y}} = \mathbf{Y} - \bar{\mathbf{y}}\mathbf{1}^\top$ where $\bar{\mathbf{y}} = \frac{1}{T+1}\sum_{t=1}^{T+1} \mathbf{y}_t$
2. Form snapshot matrices:
$$\mathbf{Y} = [\widetilde{\mathbf{y}}_1, \ldots, \widetilde{\mathbf{y}}_T], \quad \mathbf{Y}' = [\widetilde{\mathbf{y}}_2, \ldots, \widetilde{\mathbf{y}}_{T+1}]$$
3. Compute truncated SVD of $\mathbf{Y}$:
$$\mathbf{Y} \approx \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^\top$$
where $\mathbf{U} \in \mathbb{R}^{M \times N}$, $\mathbf{\Sigma} \in \mathbb{R}^{N \times N}$, $\mathbf{V} \in \mathbb{R}^{T \times N}$
4. Compute reduced-rank VAR operator:
$$\widehat{\mathbf{B}} = \mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{U}^\top$$
5. Project to reduced space and compute eigendecomposition:
$$\widetilde{\mathbf{A}} = \mathbf{U}^\top\,\mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}, \quad \widetilde{\mathbf{A}}\,\mathbf{W} = \mathbf{W}\,\mathbf{\Lambda}$$
6. Compute DMD modes:
$$\mathbf{\Phi} = \mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{W}$$

**Output:** $\widehat{\mathbf{B}}$, $\mathbf{\Phi}$, $\mathbf{\Lambda}$

## 2.2 Ramifications

Define the rank-$N$ DMD operator $\widetilde{\mathbf{B}} := \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^{+}$. By Proposition 3, $\widehat{\mathbf{B}}\mathbf{P}_{\mathbf{\Phi}} = \widetilde{\mathbf{B}}$. From now on we work with $\widetilde{\mathbf{B}}$, which has the same action as $\widehat{\mathbf{B}}$ on the rank-$N$ subspace spanned by the DMD modes.

**Proposition 4.** *Define the modal coordinates* $\widetilde{\mathbf{x}}_t := \mathbf{\Phi}^{+}\,\mathbf{y}_t$.[5] *Then the reduced-rank VAR has the linear state-space representation*

$$\begin{aligned} \widetilde{\mathbf{x}}_t &= \mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1} + \mathbf{\Phi}^{+}\widehat{\mathbf{a}}_t, \\ \mathbf{y}_t &= \mathbf{\Phi}\mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1} + \widehat{\mathbf{a}}_t, \end{aligned} \tag{14}$$

*with* $\sum_{t=2}^{T+1} \widehat{\mathbf{a}}_t\,\mathbf{y}_{t-1}^{\top} = \mathbf{0}$ *and* $\widehat{\mathbf{\Omega}} = T^{-1}\sum_{t=2}^{T+1} \widehat{\mathbf{a}}_t\widehat{\mathbf{a}}_t^{\top}$.

*Proof.* From $\widetilde{\mathbf{B}} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^{+}$, we have $\mathbf{y}_t = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^{+}\,\mathbf{y}_{t-1} + \widehat{\mathbf{a}}_t$. Premultiplying by $\mathbf{\Phi}^{+}$ yields (14). □

The dynamic modes $\widetilde{\mathbf{x}}_t$ in Proposition 4 evidently have a moving average representation

$$\widetilde{\mathbf{x}}_{t+j} = \mathbf{\Lambda}^{j}\widetilde{\mathbf{x}}_t + \sum_{s=0}^{j-1} \mathbf{\Lambda}^{s}\mathbf{\Phi}^{+}\widehat{\mathbf{a}}_{t+j-s}. \tag{15}$$

Under serially uncorrelated residuals $\{\widehat{\mathbf{a}}_t\}$ with covariance $\widehat{\mathbf{\Omega}}$, the $j$-step-ahead conditional covariances of the modes are

$$\mathbb{E}\left[(\widetilde{\mathbf{x}}_{t+j} - \mathbb{E}[\widetilde{\mathbf{x}}_{t+j} \mid \widetilde{\mathbf{x}}_t])(\widetilde{\mathbf{x}}_{t+j} - \mathbb{E}[\widetilde{\mathbf{x}}_{t+j} \mid \widetilde{\mathbf{x}}_t])^{\top}\right] = \sum_{s=0}^{j-1} \mathbf{\Lambda}^{s}\mathbf{\Phi}^{+}\widehat{\mathbf{\Omega}}(\mathbf{\Phi}^{+})^{\top}(\mathbf{\Lambda}^{\top})^{s} \tag{16}$$

When $\mathbf{\Lambda}$ is complex-valued, $\cdot^{\top}$ denotes the conjugate transpose. See Appendix B.1 for more details.

[5] When $\mathbf{\Phi}$ has full column rank, $\widetilde{\mathbf{x}}_t$ gives the coordinates of $\mathbf{P}_{\mathbf{\Phi}}\,\mathbf{y}_t$ in the mode basis. When $\mathbf{\Phi}$ is rank-deficient, $\widetilde{\mathbf{x}}_t$ is the minimum-norm least-squares coefficient of $\mathbf{y}_t$ regressed on the columns of $\mathbf{\Phi}$.

Collectively, Propositions 1–4 establish how DMD estimates a reduced-rank VAR(1) representation of the observable process $\{\mathbf{y}_t\}$ and how the fitted reduced-rank VAR(1) can be expressed in terms of dynamic modes $\mathbf{\Phi}$ and eigenvalues $\mathbf{\Lambda}$. When $\mathbf{G}$ has full column rank and the truncation rank is chosen as $N = N_x$, the fitted realization can be related to the state dynamics up to a similarity transformation: in population, the nonzero eigenvalues in $\mathbf{\Lambda}$ match the eigenvalues of $\mathbf{A}$ (Proposition 2(ii)), and under standard eigen-gap conditions the column space of $\mathbf{\Phi}$ estimates the column space of $\mathbf{G}$. When $\mathbf{G}$ is rank-deficient, Proposition 4 still yields a valid state-space representation for the reduced-rank VAR in $\{\mathbf{y}_t\}$, but $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ should be interpreted as describing the most salient observable dynamics and do not in general identify the structural transition matrix $\mathbf{A}$ (Proposition 1).

### 2.3 Two Innovations Representations

Section 2.1 described how to use DMD to estimate a reduced-rank first-order vector autoregression and then to use it to cast representation (14) in terms of dynamic modes $\widetilde{\mathbf{x}}_t$. In this section, we rewrite system (14) in a one-step-ahead form that resembles an innovations representation for the fitted reduced-rank VAR(1). Define the one-step-ahead predictor in modal coordinates by

$$\widehat{\mathbf{x}}_t := \mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1} = \mathbf{\Lambda}\mathbf{\Phi}^{+}\,\mathbf{y}_{t-1}\,. \tag{17}$$

Here $\widehat{\mathbf{x}}_t$ is defined from the fitted VAR(1) and should not be confused with the Kalman predictor $\widehat{\mathbf{x}}_t$ in (2). From (17), $\widehat{\mathbf{x}}_t = \mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1}$ and, by shifting the index, $\widehat{\mathbf{x}}_{t+1} = \mathbf{\Lambda}\widetilde{\mathbf{x}}_t$. Substituting $\widehat{\mathbf{x}}_t = \mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1}$ into the measurement equation, i.e., the second equation of (14), gives us

$$\mathbf{y}_t = \mathbf{\Phi}\widehat{\mathbf{x}}_t + \widehat{\mathbf{a}}_t.$$

Because $\widehat{\mathbf{a}}_t$ is the least-squares residual from regressing $\mathbf{y}_t$ on $\mathbf{y}_{t-1}$ (equivalently, on $\mathbf{\Phi\Lambda}\widetilde{\mathbf{x}}_{t-1}$), it satisfies the sample normal equations $\sum_{t=2}^{T+1} \widehat{\mathbf{a}}_t \, \mathbf{y}_{t-1}^\top = \mathbf{0}$. Now time-shift the first equation of system (14) forward to obtain

$$\widetilde{\mathbf{x}}_{t+1} = \mathbf{\Lambda}\widetilde{\mathbf{x}}_t + \mathbf{\Phi}^+\widehat{\mathbf{a}}_{t+1},$$

then multiply both sides by $\mathbf{\Lambda}$ to get

$$\mathbf{\Lambda}\widetilde{\mathbf{x}}_{t+1} = \mathbf{\Lambda}^2\widetilde{\mathbf{x}}_t + \mathbf{\Lambda\Phi}^+\widehat{\mathbf{a}}_{t+1}.$$

Using $\widehat{\mathbf{x}}_{t+2} = \mathbf{\Lambda}\widetilde{\mathbf{x}}_{t+1}$ and $\widehat{\mathbf{x}}_{t+1} = \mathbf{\Lambda}\widetilde{\mathbf{x}}_t$, this becomes

$$\widehat{\mathbf{x}}_{t+2} = \mathbf{\Lambda}\widehat{\mathbf{x}}_{t+1} + \mathbf{\Lambda\Phi}^+\widehat{\mathbf{a}}_{t+1}.$$

Shifting indices back by one yields our pseudo innovations representation

$$\begin{aligned} \widehat{\mathbf{x}}_{t+1} &= \mathbf{\Lambda}\widehat{\mathbf{x}}_t + \mathbf{\Lambda\Phi}^+\widehat{\mathbf{a}}_t, \\ \mathbf{y}_t &= \mathbf{\Phi}\widehat{\mathbf{x}}_t + \widehat{\mathbf{a}}_t. \end{aligned} \tag{18}$$

Equation (18) gives a one-step-ahead (pseudo-innovations) state-space realization of the fitted rank-$N$ VAR(1): it reproduces the fitted linear predictor $\widetilde{\mathbf{B}}\,\mathbf{y}_{t-1} = \mathbf{\Phi}\widehat{\mathbf{x}}_t$ and the associated residuals $\widehat{\mathbf{a}}_t := \mathbf{y}_t - \widetilde{\mathbf{B}}\,\mathbf{y}_{t-1}$ from the reduced-rank coefficient $\widetilde{\mathbf{B}} = \mathbf{\Phi\Lambda\Phi}^+$. This construction parallels the steady-state innovations representation (2) of the baseline state-space model (1), in which $(\mathbf{A}, \mathbf{G}, \mathbf{K})$ and innovations $\mathbf{a}_t$ generate the one-step-ahead predictor and update.

The difference is informational: $\widehat{\mathbf{a}}_t$ are defined by the least-squares projection used to fit the VAR, so they impose only the sample orthogonality conditions $\sum_{t=2}^{T+1} \widehat{\mathbf{a}}_t \, \mathbf{y}_{t-1}^\top = \mathbf{0}$ (and, in population, $\mathbb{E}[\widehat{\mathbf{a}}_t \, \mathbf{y}_{t-1}^\top] = \mathbf{0}$), whereas the innovations $\mathbf{a}_t$ in (2) satisfy $\mathbf{a}_t \perp \mathbf{y}^{t-1}$. When the data-generating process is VAR(1) with white innovations (for example,

under the injective-measurement case of Proposition 2(i)), these notions coincide and (18) becomes an innovations representation. This discussion preludes the information-projection interpretation of DMD in Section 3.

Motivated by this analogy, a realization of the fitted VAR(1) is obtained by setting

$$\widehat{\mathbf{A}} := \mathbf{\Lambda}, \quad \widehat{\mathbf{K}} := \mathbf{\Lambda}\mathbf{\Phi}^{+}, \quad \widehat{\mathbf{G}} := \mathbf{\Phi}. \tag{19}$$

## 3 Approximating Versus Underlying Structural Model

In the spirit of Koopmans (1947) and Sargent and Sims (1977), we regard the model in Section 2 as a Kepler-stage descriptive model whose role is to detect and organize data patterns that a Newton-stage structural model should be designed to interpret and explain. Lucas (1987) and other leading 20th-century macroeconomists viewed descriptive findings like those of Burns and Mitchell (1946) as providing empirical underpinnings for a "neo-classical synthesis" that separates macroeconomic from microeconomic analysis of redistribution and social insurance.[6] That descriptive statistical work indicates that the evolution of the macroeconomic quantities reported in National Income and Product Accounts emerged from the operation of some type of "homogenizing mechanism" that aggregates a myriad of microeconomic shocks into one or two macroeconomic "factors" and shocks.[7] According to Koopmans (1947) and Lucas (1987), the role of structural macroeconomic theory is to interpret and explain these regularities.

In this section, we formulate a particular homogenizing mechanism, namely a competitive equilibrium in the complete-markets tradition of Arrow–Debreu. We use it to generate a vector stochastic process for a cross-section of households' income and consumption, and then use the statistical model in Section 2 to describe the

[6] See Sargent (2015, 2024).

[7] Milton Friedman suggested the term "homogenizing mechanism" to Sargent in personal conversations in 1976.

simulated data. In this way, we learn how DMD-recovered parameters depend on the parameters of the underlying structural model. In Sargent et al. (2025), we use our Section 2 model to study the dynamics of CEX cross-sections of US personal income and consumption, so here we take an off-the-shelf Arrow–Debreu model, namely, a member of a class of readily computable linear-quadratic structures presented by Hansen and Sargent (2013). Thus, in the spirit of White (1982) and Hansen and Sargent (1993), our purpose is to study properties of our Section 2 descriptive model as a presumably misspecified approximation to a particular structural model. We do this by constructing an "information projection" of the structural model onto the DMD model.[8]

In Sargent et al. (2025), we detected evidence of substantial redistribution and insurance in US CEX data from 1990 to 2023. The structural model presented in this section delivers such outcomes. It is a linear-quadratic DSGE economy with heterogeneous households whose preferences satisfy Gorman aggregation conditions (Hansen and Sargent, 2013, ch. 12). Hence, competitive equilibrium prices and aggregate quantities can be computed without tracking the cross-sectional wealth distribution and household-level consumption allocations respond to aggregate dynamics through time-invariant Gorman weights.

In our structural model, two "homogenizing mechanisms"' are active, one exogenous, the other endogenous. We build in the exogenous part of the mechanism when we specify that households' exogenous stochastic consumption endowments, i.e., their "Lucas trees", share a common dynamic factor.[9] The endogenous part comes from the extensive risk-sharing that competitive and complete markets deliver. As we shall see, it is easy for us to simulate efficient redistributions by simply tilting the vector of Pareto weights that emerge from an initial competitive equilibrium allocation. We take

[8] See footnote 13 below for an explanation of "information projection". See Sargent (1976) for an early application.

[9] Eberly and Wang (2025) provide the canonical "two-tree" model that extends the classic one-tree model of Lucas (1978). Our Hansen and Sargent (2013, ch. 12) model has as many trees as households.

advantage of that feature to generate outcomes that resemble ones uncovered by our statistical analysis with the CEX data in Sargent et al. (2025), as exhibited by comparing Figure 3 below with Figure 3 in Sargent et al. (2025). Among other things, we also study how well, when applied to a high-dimensional household panel, DMD recovers low-dimensional aggregate dynamics that drive household incomes and consumptions in the DSGE model that actually generates the data in our system of artificial households and traders.

### 3.1 Gorman Aggregation Environment

Time is discrete, $t \in \{0, 1, 2, \ldots\}$. There are $J$ households, indexed by $j$, who share a common discount factor $\beta \in (0, 1)$ and information set, but differ in preferences and endowments. Households consume a single final good and supply a single intermediate input, which for convenience we sometimes call "labor". Household $j$ chooses $\{c_{jt}, \ell_{jt}\}_{t \geq 0}$ to maximize

$$-\frac{1}{2}\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[(s_{jt} - b_{jt})^\top (s_{jt} - b_{jt}) + \ell_{jt}^\top \ell_{jt}\right] \tag{20}$$

subject to a household service technology

$$\begin{aligned} s_{jt} &= \Lambda_s h_{j,t-1} + \Pi_s c_{jt}, \\ h_{jt} &= \Delta_h h_{j,t-1} + \Theta_h c_{jt}, \end{aligned}$$

and an Arrow–Debreu time-0 intertemporal budget constraint

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t p_{0t} \cdot c_{jt} = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t (w_{0t} \ell_{jt} + \alpha_{0t} \cdot d_{jt}) + v_0 \cdot k_{j,-1},$$

Here $c_{jt}$ is consumption, $\ell_{jt}$ is labor supply, $h_{jt}$ is a household durable stock, and $s_{jt}$ is the associated service flow. $p_{0t}$ is the price of consumption, $w_{0t}$ is the price of the

intermediate input, $\alpha_{0t}$ is the price vector of endowment goods, and $v_0$ is the price of the initial capital stock $k_{j,-1}$. Bliss points $b_{jt} = \mathbf{U}_b^j \mathbf{z}_t$ and endowments $d_{jt} = \mathbf{U}_d^j \mathbf{z}_t$ are linear in an exogenous state vector $\mathbf{z}_t \in \mathbb{R}^{n_z}$ that follows

$$\mathbf{z}_{t+1} = \mathbf{A}_{22}\, \mathbf{z}_t + \mathbf{C}_2\, \mathbf{w}_{t+1}, \quad \mathbf{w}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{21}$$

Heterogeneity enters only through the loadings $(\mathbf{U}_b^j, \mathbf{U}_d^j)$ and initial stocks $(h_{j,-1}, k_{j,-1})$; the technology parameters $(\Lambda_s, \Pi_s, \Delta_h, \Theta_h)$ are common across households. This structure implies individual demands are affine in wealth with common marginal propensities, a necessary condition for Gorman aggregation (Gorman, 1953).

### 3.2 Aggregate Dynamics and the Sharing Rule

Competitive equilibrium allocations are Pareto efficient and can be computed from a representative-agent planning problem. Define aggregates $c_t := \sum_j c_{jt}$, $h_t := \sum_j h_{jt}$, $k_t := \sum_j k_{jt}$, and let $\mathbf{U}_b := \sum_j \mathbf{U}_b^j$, $\mathbf{U}_d := \sum_j \mathbf{U}_d^j$ so that $b_t = \mathbf{U}_b\, \mathbf{z}_t$ and $d_t = \mathbf{U}_d\, \mathbf{z}_t$. The planner maximizes

$$-\frac{1}{2}\, \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[ (s_t - b_t)^\top (s_t - b_t) + g_t^\top g_t \right] \tag{22}$$

subject to

$$\Psi_c c_t + \Psi_g g_t + \Psi_i i_t = \Gamma k_{t-1} + d_t, \quad k_t = \Delta_k k_{t-1} + \Theta_k i_t,$$
$$h_t = \Delta_h h_{t-1} + \Theta_h c_t, \quad s_t = \Lambda_s h_{t-1} + \Pi_s c_t.$$

In (22), $g_t$ is an aggregate intermediate good and $i_t$ is investment. Let $\mu_{0j}^w$ denote household $j$'s time-zero marginal utility of wealth and define the Pareto weight associated with the competitive equilibrium for an economy with initial stocks $(k_{j,-1}, h_{j,-1})$ and

exogenous endowments $\{d_{j,t}\}$ across agents $j$ as

$$\mu_j := \frac{\mu_{0j}^w}{\sum_{i=1}^J \mu_{0i}^w}. \tag{23}$$

The consumption allocation rule takes the form

$$c_{jt} = \mu_j c_t + \tilde{\chi}_{jt}, \tag{24}$$

where $\mu_j c_t$ is household $j$'s proportional share of aggregate consumption and $\tilde{\chi}_{jt}$ is a deviation term determined by preference heterogeneity. The deviations sum to zero: $\sum_j \tilde{\chi}_{jt} = 0$. The weight $\mu_j$ depends only on time-0 conditions and is time-invariant. Holding aggregate endowments and initial aggregate stocks fixed, if we assume a new set of Pareto weights that sum to 1, then (24) describes the new allocation while leaving aggregate dynamics unchanged. In this way, we can describe the outcomes of a Pareto-efficient redistribution scheme while leaving open details about the exact redistributions. There is an equivalence class of distributions of initial capital stocks and endowments that validate a given set of Pareto weights. We therefore model redistribution by directly perturbing the Pareto weights.

For a special case of our model in which preference shocks are shut down, an Arrow–Debreu complete-markets allocation can be implemented with a limited set of markets: a mutual fund holding all endowment claims and a one-period riskless bond. When the riskless return is constant and the deviation baseline $\{\tilde{\chi}_{jt}\}$ is known at time 0, all households hold the same portfolio of risky assets (the mutual fund), and dynamic rebalancing occurs solely through the bond market. This trading structure implements an efficient allocation and extends a two-fund theorem of Rubinstein (1974) to a multiperiod setting.[10]

[10] Sargent and Stachurski (2025) provides a more detailed discussion of the model in their lecture.

### 3.3 A 100-Household Example

We now construct a large-panel experiment in which DMD is applied to a high-dimensional cross-section of household outcomes generated by the Gorman economy. Set $J = 100$ and define vectors $\boldsymbol{\eta}_t := (\eta_{J_a+1,t}, \ldots, \eta_{J,t})'$ as the idiosyncratic endowment states for $j > J_a$ and $\boldsymbol{\xi}_t := (\xi_{1,t}, \ldots, \xi_{J,t})'$ as the preference-shock states. The exogenous state is

$$\mathbf{z}_t = \begin{bmatrix} 1 \\ d_{a,t} \\ \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{bmatrix}.$$

The aggregate component follows the AR(1) process

$$d_{a,t+1} = \rho_1 d_{a,t} + \sigma_a \boldsymbol{\varepsilon}_{a,t+1}, \qquad (\rho_1, \sigma_a) = (0.95, 0.5).$$

where $\boldsymbol{\varepsilon}_{a,t+1} \sim \mathcal{N}(0,1)$ is the aggregate innovation (a component of $\mathbf{w}_{t+1}$ in (21)). We set $J_a = 50$ and construct endowments so that idiosyncratic shocks cancel in the aggregate:

$$\begin{aligned} d_{jt} &= \alpha_j + \phi_j d_{a,t} + \eta_{j,t}, \quad j = J_a + 1, \ldots, J, \\ d_{jt} &= \alpha_j + \phi_j d_{a,t} - \frac{1}{J_a} \sum_{k=J_a+1}^{J} \eta_{k,t}, \quad j = 1, \ldots, J_a. \end{aligned}$$

In this setting, $\sum_{j=1}^{J} d_{jt}$ depends on $\{d_{a,t}\}$ but not on $\{\eta_{j,t}\}$ once we normalize $\sum_{j=1}^{J} \phi_j = 1$. We draw $\alpha_j \sim \mathcal{U}[3,5]$ and $\tilde{\phi}_j \sim \mathcal{U}[0.5, 1.5]$, then set $\phi_j = \tilde{\phi}_j / \sum_{i=1}^{J} \tilde{\phi}_i$. To generate an economy where low-income households experience more idiosyncratic risk, we define $p_j$ as one minus the percentile rank of $\alpha_j$ in the cross-section and set, for $j > J_a$,

$$\sigma_j = 0.2 + (5.0 - 0.2)\, p_j^2, \qquad \rho_j^d = 0.0 + (0.90 - 0.0)\, p_j,$$

in the AR(1) process $\eta_{j,t+1} = \rho_j^d \eta_{j,t} + \sigma_j \varepsilon_{j,t+1}^d$. The innovations $\{\varepsilon_{j,t+1}^d\}$ are i.i.d. $\mathcal{N}(0,1)$ across $j$ and $t$ and are components of $\mathbf{w}_{t+1}$. Preference shocks are muted in this experiment by setting the innovation loadings on $\xi_{j,t}$ to zero. We interpret changes in Pareto weights as an incompletely described, Pareto-efficient tax-and-transfer scheme that redistributes resources across households without changing aggregate dynamics. To implement redistribution, we compute an alternative Pareto-efficient allocation that replaces the competitive equilibrium weights $\{\mu_j\}$ with redistributed weights $\{\mu_j^*\}$ defined by

$$\mu_j^* = (1-a)\mu_j + a/J, \qquad a = 0.8,$$

so that $\sum_{j=1}^J \mu_j^* = 1$. As mentioned above, there is an equivalence class of efficient tax-and-transfer systems that imply these Pareto weights.

We simulate the economy for 2,000 periods and discard the first 200 as burn-in. Figure 1 plots simulated household consumption and post-trade dividend-income panels, and Figure 2 summarizes the redistribution in Pareto weights. Under the sharing rule (24), the aggregate component of consumption is $\mu_j c_t$, the source of strong comovements across households (Figure 1).

Figure 2 shows that the reduced-form redistribution tilts weights away from those implied by initial marginal utilities of wealth toward a more uniform assignment. Because Gorman aggregation pins down the aggregate allocation independently of the Pareto weights, this redistribution operates only through the sharing rule: it reshuffles the cross-sectional allocation while leaving the aggregate dynamics $(c_t, k_t)$ unchanged.

To visualize and model the tax-and-transfer mechanism in the economy, we define the household income process and study how redistribution affects household income and consumption. Let $y_{jt}(\boldsymbol{\omega})$ be household $j$'s net income for a given Pareto-weight vector $\boldsymbol{\omega}$, and define household $j$'s net income at time $t$ as

$$y_{jt}(\boldsymbol{\omega}) = \omega_j d_t + (R-1) w_{j,t-1}, \tag{25}$$

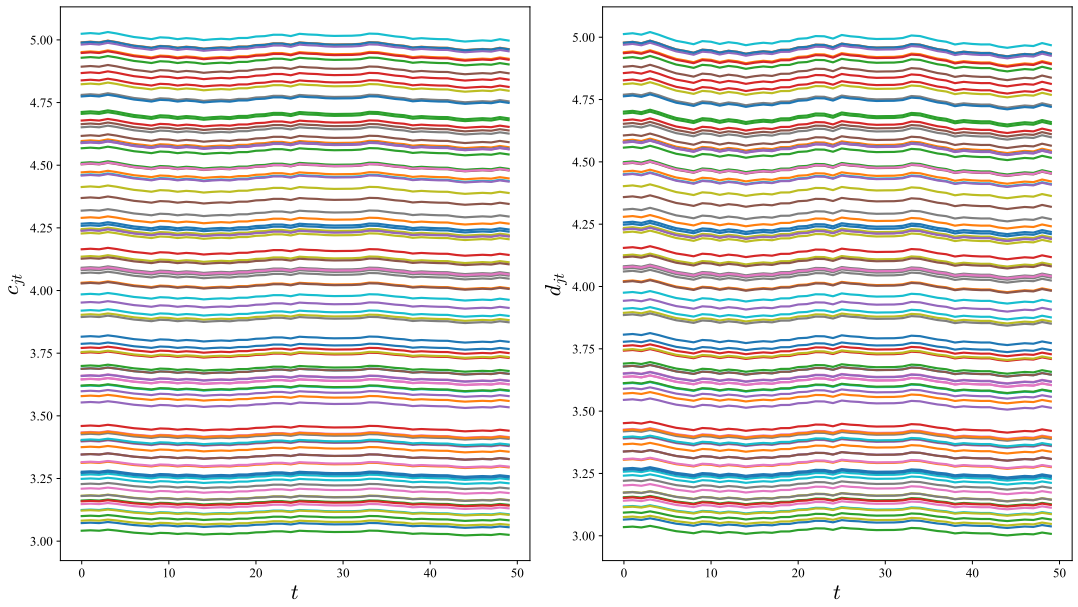


**Figure 1:** Household consumption and post-trade dividend-income paths in the 100-household economy (after burn-in). Each line corresponds to a household.

where $d_t$ is the aggregate endowment, $R = 1/\beta$ is the gross return, and $w_{j,t-1} = \omega_j k_{t-1} + \hat{k}_{j,t-1}$ is household $j$'s total asset comprising a proportional capital share $\omega_j k_{t-1}$ and a bond position $\hat{k}_{j,t-1}$. The first term is dividend income from holding $\omega_j$ shares of the aggregate endowment; the second term is the net return on wealth. Let $y_{jt}^{pre} := y_{jt}(\boldsymbol{\mu})$ denote household $j$'s net income under the competitive equilibrium weight vector $\{\mu_j\}$, let $y_{jt}^{post} := y_{jt}(\boldsymbol{\mu}^*)$ denote net income under the redistributed weights $\{\mu_j^*\}$, and let $c_{jt}^{post}$ denote post-redistribution consumption.

The effect of redistribution on household income and consumption is clearly visible in Figure 3, which plots cross-sectional percentiles of pre- and post-redistribution income and post-redistribution consumption.

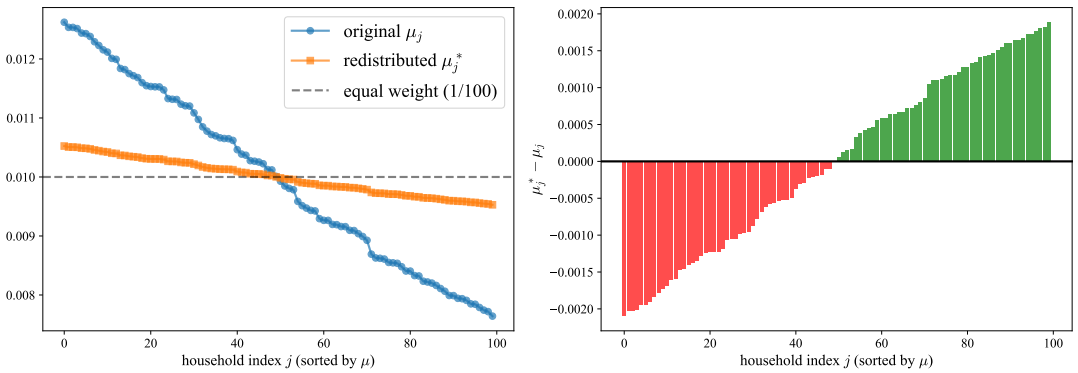


**Figure 2:** Pareto weights in the 100-household economy. The left panel plots the Gorman weights $\{\mu_j\}$ and redistributed weights $\{\mu_j^*\}$ after sorting households by $\mu_j$. The right panel plots $\mu_j^* - \mu_j$.

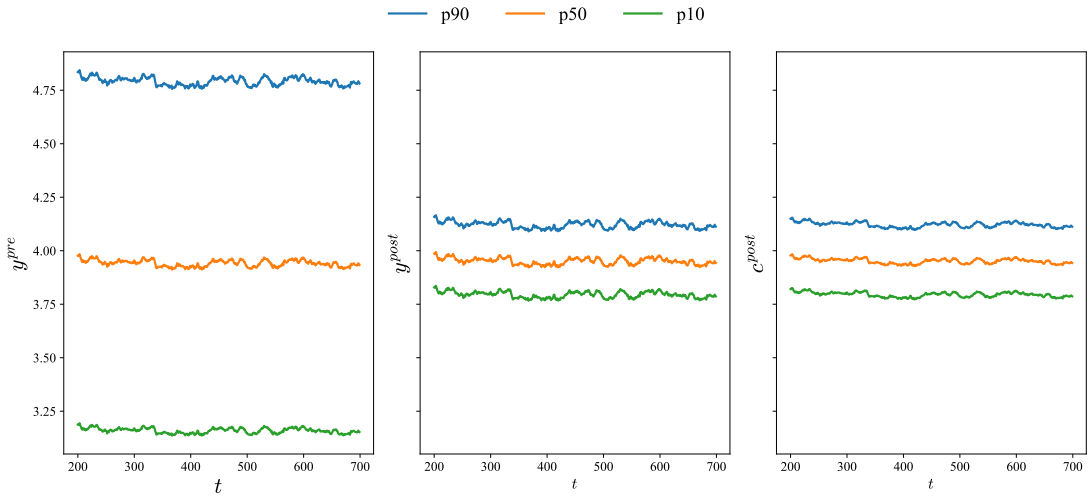


**Figure 3:** Cross-sectional percentiles (p10, p50, p90) of pre-redistribution income $y_{jt}^{pre}$, post-redistribution income $y_{jt}^{post}$, and post-redistribution consumption $c_{jt}^{post}$ in the 100-household experiment (after burn-in).

### 3.4 Ground Truth State-Space Representation

A solution to the planner's problem yields equilibrium prices and allocations as linear functions of the aggregate state. This provides a ground-truth state-space representation for the high-dimensional household panel. The aggregate state vector is

$$\mathbf{x}_t = \begin{bmatrix} h_{t-1} \\ k_{t-1} \\ \mathbf{z}_t \end{bmatrix},$$

where $h_{t-1}$ is the aggregate household durable stock, $k_{t-1}$ is the aggregate capital stock. The equilibrium law of motion for the state evolves according to

$$\mathbf{x}_{t+1} = \mathbf{A}_o \, \mathbf{x}_t + \mathbf{C}_o \, \mathbf{w}_{t+1}, \qquad \mathbf{w}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{26}$$

Any equilibrium quantity is a linear function of the state. Stacking aggregate consumption, government spending, investment, household durables, capital, endowments, bliss points, and services gives a measurement equation

$$\mathbf{y}_t^{\text{agg}} := \begin{bmatrix} c_t \\ g_t \\ i_t \\ h_t \\ k_t \\ d_t \\ b_t \\ s_t \end{bmatrix} = \mathbf{G}_o \, \mathbf{x}_t, \tag{27}$$

where $\mathbf{G}_o$ stacks the pertinent selection vectors.

Suppose an econometrician observes only household-level income and consumption data, without direct knowledge of the underlying state $\mathbf{x}_t$ or the structural parameters governing its evolution. That is, (26) is out of reach. The econometrician observes a high-dimensional panel that stacks households' pre- and post-redistribution incomes and post-redistribution consumptions as

$$\mathbf{y}_t^{\mathrm{GL}} := \begin{bmatrix} y_{1t}^{pre} \\ \vdots \\ y_{Jt}^{pre} \\ y_{1t}^{post} \\ \vdots \\ y_{Jt}^{post} \\ c_{1t}^{post} \\ \vdots \\ c_{Jt}^{post} \end{bmatrix} \in \mathbb{R}^M, \qquad M = 3J.$$

Define the $J$-vectors $\mathbf{y}_t^{pre} := (y_{1t}^{pre}, \ldots, y_{Jt}^{pre})^\top$, $\mathbf{y}_t^{post} := (y_{1t}^{post}, \ldots, y_{Jt}^{post})^\top$, and $\mathbf{c}_t^{post} := (c_{1t}^{post}, \ldots, c_{Jt}^{post})^\top$ so that $\mathbf{y}_t^{\mathrm{GL}} = [\mathbf{y}_t^{pre}; \mathbf{y}_t^{post}; \mathbf{c}_t^{post}]$. Since (as verified below) each component of $\mathbf{y}_t^{\mathrm{GL}}$ is well approximated as a linear function of the state, the household panel is well approximated by a state-space representation of the form

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}_o\, \mathbf{x}_t + \mathbf{C}_o\, \mathbf{w}_{t+1}, \\ \mathbf{y}_t^{\mathrm{GL}} &= \mathbf{G}^{\mathrm{GL}}\, \mathbf{x}_t, \end{aligned} \tag{28}$$

which is a special case of (1) with $(\mathbf{A}, \mathbf{C}, \mathbf{G}) = (\mathbf{A}_o, \mathbf{C}_o, \mathbf{G}^{\mathrm{GL}})$. The measurement matrix

$\mathbf{G}^{\mathrm{GL}}$ has the block structure

$$\mathbf{G}^{\mathrm{GL}} = \begin{bmatrix} \mathbf{G}^{y,\mathrm{pre}} \\ \mathbf{G}^{y,\mathrm{post}} \\ \mathbf{G}^{c,\mathrm{post}} \end{bmatrix} \qquad \mathbf{G}^{\mathrm{GL}} \in \mathbb{R}^{M \times N_x},$$

where $N_x = 154$ is the dimension of $\mathbf{x}_t$.[11] and $\mathbf{G}^{y,\mathrm{pre}}, \mathbf{G}^{y,\mathrm{post}}, \mathbf{G}^{c,\mathrm{post}} \in \mathbb{R}^{J \times N_x}$ map $\mathbf{x}_t$ to $\mathbf{y}_t^{pre}$, $\mathbf{y}_t^{post}$, and $\mathbf{c}_t^{post}$, respectively. To write $\mathbf{G}^{\mathrm{GL}}$ in closed form, we express each observable as a linear function of the aggregate state $\mathbf{x}_t$.

For income, combining (25) with the asset decomposition $w_{j,t-1} = \omega_j k_{t-1} + \hat{k}_{j,t-1}$ gives

$$y_{jt}(\boldsymbol{\omega}) = \omega_j d_t + (R-1)\omega_j k_{t-1} + (R-1)\hat{k}_{j,t-1}, \tag{29}$$

where the first two terms depend only on aggregates, but the third term involves household $j$'s idiosyncratic bond position $\hat{k}_{j,t-1}$. In the limited-markets implementation of the complete-markets allocation, the bond position is driven by the sharing-rule deviation through the law of motion

$$\hat{k}_{jt} = R\,\hat{k}_{j,t-1} - \tilde{\chi}_{jt}, \tag{30}$$

where $\tilde{\chi}_{jt}$ is the deviation term from the sharing rule (24). The initial bond position is pinned down by the present value of future deviations:

$$\hat{k}_{j0} = \sum_{s=1}^{\infty} \beta^s \tilde{\chi}_{js}. \tag{31}$$

In general, writing $y_{jt}$ as a static function of $\mathbf{x}_t$ alone requires augmenting the state with $\{\hat{k}_{j,t-1}\}_{j=1}^{J}$. However, when preference shocks are muted so that $\tilde{\chi}_{jt}$ is negligible,

[11] The 154 states consist of $h_{t-1}$, $k_{t-1}$, $d_{a,t}$, $d_{a,t-1}$, $\boldsymbol{\eta}_t \in \mathbb{R}^{J-J_a}$ with $J - J_a = 50$, and $\boldsymbol{\xi}_t \in \mathbb{R}^{100}$.

equations (30)–(31) imply $\hat{k}_{jt} \approx 0$ for all $j$ and $t$, and income is well approximated by

$$y_{jt}(\boldsymbol{\omega}) \approx \omega_j \, \mathbf{S}_d \, \mathbf{x}_t + (R-1)\omega_j \, \mathbf{S}_k \, \mathbf{x}_t, \tag{32}$$

where $\mathbf{S}_d$ and $\mathbf{S}_k$ are selection vectors such that $d_t = \mathbf{S}_d \, \mathbf{x}_t$ and $k_{t-1} = \mathbf{S}_k \, \mathbf{x}_t$. This yields the income row $\mathbf{G}_j^y \approx \omega_j \mathbf{S}_d + (R-1)\omega_j \mathbf{S}_k$ with $\mathbf{G}^y \in \{\mathbf{G}^{y,\text{pre}}, \mathbf{G}^{y,\text{post}}\}$ and $\omega_j \in \{\mu_j, \mu_j^*\}$ correspondingly.

The sharing rule (24) implies

$$c_{jt} = \mu_j c_t + \tilde{\chi}_{jt}, \tag{33}$$

where the deviation $\tilde{\chi}_{jt}$ captures preference heterogeneity. Under the canonical household technology, $\tilde{\chi}_{jt}$ is driven by the deviation bliss point $\tilde{b}_{jt} := b_{jt} - \mu_j b_t$. In our calibration with $\Lambda_s = 0$, the inverse representation simplifies to

$$\tilde{\chi}_{jt} = \Pi_s^{-1} \tilde{b}_{jt}, \qquad \tilde{b}_{jt} = \mathbf{U}_b^j \, \mathbf{z}_t - \mu_j b_t, \tag{34}$$

where $\Pi_s$ is the service–consumption loading in the household technology and $b_t = \mathbf{S}_b \, \mathbf{x}_t$ is the aggregate bliss point. Substituting yields

$$\mathbf{G}_j^c = \mu_j \mathbf{S}_c + \Pi_s^{-1} \big( \mathbf{U}_b^j \, \mathbf{S}_z - \mu_j \mathbf{S}_b \big), \tag{35}$$

where $\mathbf{S}_c$ and $\mathbf{S}_z$ select aggregate consumption and the exogenous block from $\mathbf{x}_t$, respectively. For post-redistribution consumption, replace $\mu_j$ with $\mu_j^*$ in (35), i.e. $\mathbf{G}_j^{c,\text{post}} = \mu_j^* \mathbf{S}_c + \Pi_s^{-1} (\mathbf{U}_b^j \, \mathbf{S}_z - \mu_j^* \mathbf{S}_b)$.

In our numerical experiments, we mute the idiosyncratic preference shocks by setting their innovation variances to zero. This has two consequences. First, the consumption deviation $\tilde{\chi}_{jt}$ in (34) becomes nearly deterministic. Second, the bond

dynamics (30)–(31) nearly collapse: when $\tilde{\chi}_{jt} \approx 0$, the initial condition (31) gives $\hat{k}_{j0} \approx 0$, and the law of motion (30) implies $\hat{k}_{jt} \approx 0$ for all $t$. Therefore, the idiosyncratic term $(R-1)\hat{k}_{j,t-1}$ becomes negligible in (29), and approximation (32) is accurate.

Moreover, we work with a $T = 250$ subsample of the post-burn-in data. Here, $M = 300$ and $T = 250$, so $M > T$.[12] The household panel $\{\mathbf{y}_t^{\mathrm{GL}}\}_{t=1}^{T}$ is our input to the DMD procedure of Section 2.1 to estimate a reduced-rank VAR operator $\widehat{\mathbf{B}}$ and its eigendecomposition. In our laboratory, we generate aggregates from the linear state-space system (26)–(27), but the econometrician observes only microdata at the household level $\{\mathbf{y}_t^{\mathrm{GL}}\}_{t=1}^{T}$. We assess DMD's performance through diagnostics that compare estimated objects to their ground-truth counterparts. We proceed from rank selection to spectral and state recovery, then turn to cross-sectional loadings and forecasting.

Before examining empirical results, it is useful to record the population object that DMD targets. Since our DMD implementation is applied to demeaned snapshots, we work with the centered panel $\mathbf{y}_t := \mathbf{y}_t^{\mathrm{GL}} - \bar{\mathbf{y}}$, where $\bar{\mathbf{y}}$ is the sample mean of $\{\mathbf{y}_t^{\mathrm{GL}}\}_{t=1}^{T}$. The centered panel $\mathbf{y}_t$ admits a state-space representation that parallels (1)–(2). Let $(\mathbf{A}, \mathbf{C}, \mathbf{G})$ denote the state transition, shock loading, and measurement matrices of the centered state-space system (i.e., with the constant state component removed from $(\mathbf{A}_o, \mathbf{C}_o, \mathbf{G}^{\mathrm{GL}})$). The centered state-space representation is

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}\,\mathbf{x}_t + \mathbf{C}\,\mathbf{w}_{t+1}, \\ \mathbf{y}_t &= \mathbf{G}\,\mathbf{x}_t, \end{aligned} \tag{36}$$

[12] We simulate 2,000 periods, discard the first 200 periods as burn-in, and extract a 250-period subsample for DMD analysis.

where $\mathbf{w}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The associated steady-state innovations representation is

$$\begin{aligned} \widehat{\mathbf{x}}_{t+1} &= \mathbf{A}\,\widehat{\mathbf{x}}_t + \mathbf{K}_{\mathrm{GL}}\,\mathbf{a}_t, \\ \mathbf{y}_t &= \mathbf{G}\,\widehat{\mathbf{x}}_t + \mathbf{a}_t, \end{aligned} \tag{37}$$

where $\widehat{\mathbf{x}}_t := \mathbb{E}[\mathbf{x}_t \mid \mathbf{y}^{t-1}]$ is the one-step-ahead state prediction, $\mathbf{a}_t := \mathbf{y}_t - \mathbf{G}\,\widehat{\mathbf{x}}_t$ is the innovation, and the steady-state Kalman gain is

$$\mathbf{K}_{\mathrm{GL}} := \mathbf{A}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top(\mathbf{G}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top)^+, \tag{38}$$

with $\mathbf{\Sigma}_\infty := \mathbb{E}[(\mathbf{x}_t - \widehat{\mathbf{x}}_t)(\mathbf{x}_t - \widehat{\mathbf{x}}_t)^\top]$ denoting the steady-state prediction error covariance.

To derive the vector autoregressive representation, iterate (37) forward. From the first equation, $\widehat{\mathbf{x}}_t = \mathbf{A}\,\widehat{\mathbf{x}}_{t-1} + \mathbf{K}_{\mathrm{GL}}\,\mathbf{a}_{t-1}$. Substituting into the second equation of (37) gives

$$\mathbf{y}_t = \mathbf{G}\,\mathbf{A}\,\widehat{\mathbf{x}}_{t-1} + \mathbf{G}\,\mathbf{K}_{\mathrm{GL}}\,\mathbf{a}_{t-1} + \mathbf{a}_t\,. \tag{39}$$

From the second equation of (37) at $t-1$, we have $\mathbf{G}\,\widehat{\mathbf{x}}_{t-1} = \mathbf{y}_{t-1} - \mathbf{a}_{t-1}$. Repeated substitution yields

$$\mathbf{y}_t = \sum_{j=1}^{\infty} \mathbf{G}(\mathbf{A} - \mathbf{K}_{\mathrm{GL}}\,\mathbf{G})^{j-1}\,\mathbf{K}_{\mathrm{GL}}\,\mathbf{y}_{t-j} + \mathbf{a}_t\,. \tag{40}$$

Define the VAR coefficient matrices

$$\mathbf{B}_j := \mathbf{G}(\mathbf{A} - \mathbf{K}_{\mathrm{GL}}\,\mathbf{G})^{j-1}\,\mathbf{K}_{\mathrm{GL}}, \quad j \geq 1 \tag{41}$$

Then the infinite-order VAR representation becomes

$$\mathbf{y}_t = \sum_{j=1}^{\infty} \mathbf{B}_j\,\mathbf{y}_{t-j} + \mathbf{a}_t\,. \tag{42}$$

This is a vector autoregressive representation for the centered household panel $\{\mathbf{y}_t\}$.

Let $\mathbf{\Sigma}_x := \mathrm{Cov}(\mathbf{x}_t)$ denote the stationary covariance of the centered state. The population VAR(1) coefficient is the linear projection of $\mathbf{y}_t$ on $\mathbf{y}_{t-1}$:

$$\mathbf{B}^* := \mathrm{Cov}(\mathbf{y}_t, \mathbf{y}_{t-1})\ \mathrm{Cov}(\mathbf{y}_{t-1})^+ = \mathbf{G}\,\mathbf{A}\,\mathbf{\Sigma}_x\,\mathbf{G}^\top \left(\mathbf{G}\,\mathbf{\Sigma}_x\,\mathbf{G}^\top\right)^+. \tag{43}$$

When $\mathbf{B}^*$ has (effective) rank at most $N$, a rank-$N$ DMD estimator targets $\mathbf{B}^*$. We now examine the eigenvalue structure of $\mathbf{B}^*$ and $\{\mathbf{B}_j\}$ to assess the suitability of a low-rank approximation.

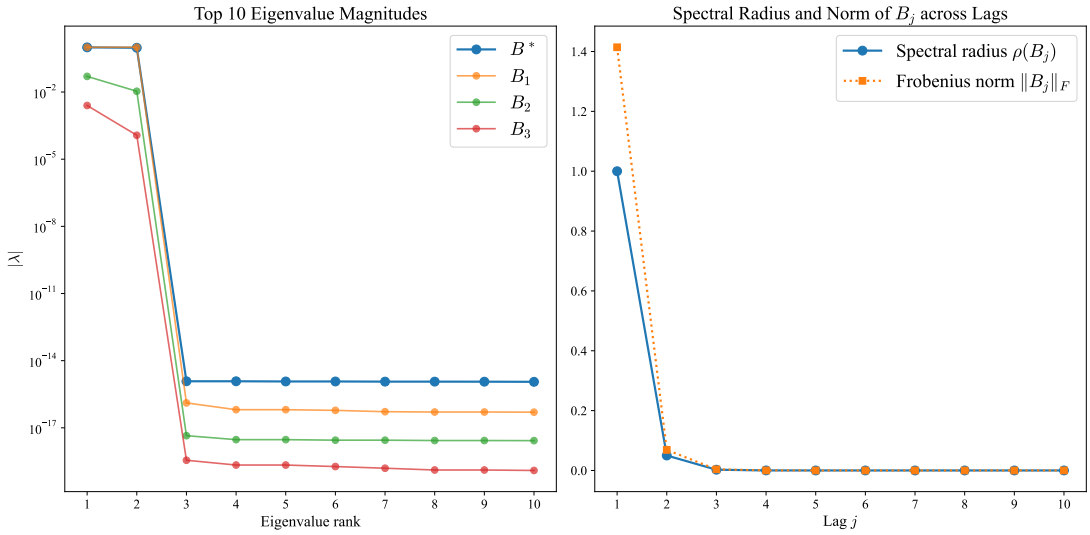


**Figure 4:** Eigenvalue structure of population VAR coefficient matrices. Left: Eigenvalue magnitudes for the population VAR(1) coefficient $\mathbf{B}^*$ and the first three infinite-order VAR coefficients $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$. Right: Spectral radius and Frobenius norm of $\mathbf{B}_j$ across lags.

Figure 4 provides evidence that the population first-order VAR coefficient $\mathbf{B}^*$ and the infinite-order VAR coefficients $\{\mathbf{B}_j\}$ are both effectively low-rank. The left panel plots the eigenvalue magnitudes of $\mathbf{B}^*$ and the first three infinite-order VAR coefficients $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$ from (42). For each matrix, the top two eigenvalues dominate,

with magnitudes around $10^{-1}$ to 1, while all other eigenvalues are numerically 0.

The nearly rank-2 structure of the VAR coefficient matrices is inherited from the economic structure of our DSGE model. Under the sharing rule (24), household consumption $c_{jt} = \mu_j c_t + \tilde{\chi}_{jt}$ is dominated by the aggregate component $\mu_j c_t$, which under the planner's linear decision rule is itself a linear function of $(k_{t-1}, d_{a,t})$. Similarly, from the income definition (25), household income $y_{jt} = \omega_j d_t + (R-1)w_{j,t-1}$ depends on aggregate endowment $d_t$ (a linear function of $d_{a,t}$) and wealth $w_{j,t-1}$, which loads on aggregate capital $k_{t-1}$. Although the full state vector $\mathbf{x}_t$ contains idiosyncratic endowment states $\{\eta_{j,t}\}$ with dimension $J - J_a$, these components are constructed to cancel in the aggregate (Section 3.3) and do not contribute additional predictive directions for the centered panel. Hence, the $M$-dimensional observable vector $\mathbf{y}_t$ lies in a two-dimensional affine subspace spanned by loadings on $(k_{t-1}, d_{a,t})$, and the population VAR coefficients inherit this low-rank structure.

The right panel of Figure 4 shows the spectral radius $\rho(\mathbf{B}_j)$ and Frobenius norm $\| \mathbf{B}_j \|_F$ across lags $j = 1, \ldots, 10$. Both decay rapidly: the spectral radius falls from near unity at lag 1 to below 0.1 by lag 2, reflecting the geometric decay rate governed by eigenvalues of the closed-loop observer $\mathbf{A} - \mathbf{K}_{\text{GL}}\,\mathbf{G}$. This decay implies that even though the true representation (42) is an infinite-order VAR, most predictive information concentrates in the first lag, making a VAR(1) a promising approximation. Heuristically, these results justify using a small $N$ in DMD: when the spectrum of $\mathbf{B}^*$ is dominated by a few eigenvalues, a low-rank approximation captures most of the predictive content.

By viewing our estimated DMD model as an "information projection" of our structural model onto the DMD family, we can analyze properties of our descriptive Section 2 model as an approximation to our structural DSGE.[13] Let $\theta$ denote the

[13] Csiszár and Matus (2003) and Nielsen (2018) describe information projections. Let $\{f_\theta(x)\}_{\theta\in\Theta}$ and $\{g_\delta(x)\}_{\delta\in\Delta}$ be two collections (manifolds) of probability distributions for outcomes $x \in X$. When model $g_{\delta_o}(x)$ governs the data, a population maximum likelihood estimator $\theta_o$ of parameter vector $\theta \in \Theta$ of misspecified statistical model $f_\theta(x)$

parameters of our DMD model and let $\delta$ denote the parameters of our structural DSGE model. When the structural model with parameter $\delta_o$ generates the data, information projection selects the KL-minimizing DMD parameter $\theta_o$. By studying the mapping $\delta_o \mapsto \theta_o$, we can learn how the parameters of the approximating descriptive model reflect the parameters of the structural model. In the remainder of this section, we describe aspects of this information projection in our setting.

Footnote 13 specializes in our setting to a conditional KL projection. We approximate the true one-step-ahead conditional law of $\mathbf{y}_t$ by a Gaussian VAR(1) conditional family, and define the best-fitting VAR coefficient as the KL minimizer.

When $\mathbf{\Gamma}_0 := \mathbb{E}[\mathbf{y}_{t-1}\,\mathbf{y}_{t-1}^\top]$ is singular, $\mathbf{y}_t$ is supported on the subspace $\mathcal{S} := \mathrm{range}(\mathbf{\Gamma}_0) \subset \mathbb{R}^M$. Let $r := \mathrm{rank}(\mathbf{\Gamma}_0)$, choose $\mathbf{R} \in \mathbb{R}^{M\times r}$ with orthonormal columns spanning $\mathcal{S}$, and write $\tilde{\mathbf{y}}_t := \mathbf{R}^\top \mathbf{y}_t \in \mathbb{R}^r$ (so $\mathbf{y}_t = \mathbf{R}\tilde{\mathbf{y}}_t$ a.s.). Assume the conditional law of $\tilde{\mathbf{y}}_t$ given $\mathbf{y}^{t-1}$ admits a Lebesgue density $p_t(\cdot \mid \mathbf{y}^{t-1})$ on $\mathbb{R}^r$ with $\mathbb{E}\,|\log p_t(\tilde{\mathbf{y}}_t \mid \mathbf{y}^{t-1})| < \infty$. For $\widetilde{\mathbf{B}} \in \mathbb{R}^{r\times r}$ and $\widetilde{\mathbf{\Omega}} \succ \mathbf{0}$, let $q_{\widetilde{\mathbf{B}},\widetilde{\mathbf{\Omega}}}(\cdot \mid \tilde{\mathbf{y}}_{t-1})$ denote the Gaussian VAR(1) density on $\mathbb{R}^r$ corresponding to $\tilde{\mathbf{y}}_t \mid \tilde{\mathbf{y}}_{t-1} \sim \mathcal{N}(\widetilde{\mathbf{B}}\tilde{\mathbf{y}}_{t-1}, \widetilde{\mathbf{\Omega}})$. [14]

Define the expected conditional KL risk

$$\mathcal{D}(\widetilde{\mathbf{B}}, \widetilde{\mathbf{\Omega}}) := \mathbb{E}\left[\int_{\mathbb{R}^r} \log\left(\frac{p_t(\tilde{\mathbf{y}} \mid \mathbf{y}^{t-1})}{q_{\widetilde{\mathbf{B}},\widetilde{\mathbf{\Omega}}}(\tilde{\mathbf{y}} \mid \tilde{\mathbf{y}}_{t-1})}\right) p_t(\tilde{\mathbf{y}} \mid \mathbf{y}^{t-1})\, d\tilde{\mathbf{y}}\right]. \tag{44}$$

minimizes the Kullback-Leibler divergence

$$\mathrm{KL}(g_{\delta_o}, f_\theta) = \int \log\left(\frac{g_{\delta_o}(x)}{f_\theta(x)}\right) g_{\delta_o}(x)dx = -H(g_{\delta_o}) - \mathbb{E}_{g_{\delta_o}} \log f_\theta(x),$$

where $H(g_{\delta_o}) = \int \log\left(\frac{1}{g_{\delta_o}(x)}\right) g_{\delta_o}(x)dx$ is the Shannon information of nature's probability distribution $g_{\delta_o}(x)$ and $\mathbb{E}_{g_{\delta_o}}$ denotes mathematical expectation under $g_{\delta_o}(x)$. The information projection of $g_{\delta_o}(x)$ onto $\{f_\theta(x)\}_{\theta\in\Theta}$ is distribution $f_{\theta_o}(x)$ in manifold $\{f_\theta(x)\}_{\theta\in\Theta}$ that maximum likelihood selects when nature's model $g_{\delta_o}$ generates the data, i.e., $\theta_o = \arg\max_{\theta\in\Theta} \mathbb{E}_{g_{\delta_o}} \log f_\theta(x)$ is the population maximum likelihood estimator of $\theta$ when probability distribution $g_{\delta_o}$ generates the data.

[14] Defining densities on $\mathbb{R}^M$ and projecting onto a full-rank Gaussian family yields KL divergence $+\infty$ when $\mathbf{\Gamma}_0$ is singular, because the true conditional law is supported on $\mathcal{S}$ while a full-rank Gaussian does not share that support.

By iterated expectations,

$$\mathcal{D}(\widetilde{\mathbf{B}}, \widetilde{\mathbf{\Omega}}) = \underbrace{\mathbb{E}\big[\log p_t(\tilde{\mathbf{y}}_t \mid \mathbf{y}^{t-1})\big]}_{\text{does not depend on } (\widetilde{\mathbf{B}}, \widetilde{\mathbf{\Omega}})} - \mathbb{E}\big[\log q_{\widetilde{\mathbf{B}}, \widetilde{\mathbf{\Omega}}}(\tilde{\mathbf{y}}_t \mid \tilde{\mathbf{y}}_{t-1})\big], \tag{45}$$

so minimizing $\mathcal{D}$ is equivalent to maximizing the expected Gaussian conditional log-likelihood on the support. Expanding the Gaussian log-density gives

$$\mathcal{D}(\widetilde{\mathbf{B}}, \widetilde{\mathbf{\Omega}}) = \text{const} + \frac{1}{2} \log\det(\widetilde{\mathbf{\Omega}}) + \frac{1}{2}\, \mathbb{E}\Big[(\tilde{\mathbf{y}}_t - \widetilde{\mathbf{B}}\tilde{\mathbf{y}}_{t-1})^\top \widetilde{\mathbf{\Omega}}^{-1} (\tilde{\mathbf{y}}_t - \widetilde{\mathbf{B}}\tilde{\mathbf{y}}_{t-1})\Big]. \tag{46}$$

Fix $\widetilde{\mathbf{\Omega}} \succ \mathbf{0}$. The minimizer $\widetilde{\mathbf{B}}^*$ satisfies

$$\mathbb{E}\left[(\tilde{\mathbf{y}}_t - \widetilde{\mathbf{B}}^* \tilde{\mathbf{y}}_{t-1}) \tilde{\mathbf{y}}_{t-1}^\top\right] = \mathbf{0}, \tag{47}$$

hence $\widetilde{\mathbf{B}}^* = \widetilde{\mathbf{\Gamma}}_1 \widetilde{\mathbf{\Gamma}}_0^{-1}$ where $\widetilde{\mathbf{\Gamma}}_k := \mathbb{E}[\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_{t-k}^\top]$ and $\widetilde{\mathbf{\Gamma}}_0 = \mathbf{R}^\top \mathbf{\Gamma}_0 \mathbf{R} \succ \mathbf{0}$.

Mapping back to $\mathbb{R}^M$ yields

$$\mathbf{B}^* := \mathbf{R}\, \widetilde{\mathbf{B}}^*\, \mathbf{R}^\top = \mathbf{\Gamma}_1\, \mathbf{\Gamma}_0^+,$$

so $\mathbf{B}^*$ is the KL (information) projection of the true one-step-ahead conditional law on its support onto the Gaussian VAR(1) family.

### 3.5 Application of DMD on Microdata from the Gorman Model

Up to this point, we have shown that $\mathbf{B}^*$ is effectively low-rank and that rank-$N$ DMD targets $\mathbf{B}^*$. The information projection perspective also makes precise the criterion underlying this target: DMD selects the best-fitting low-rank VAR(1) approximation to the panel's true one-step-ahead conditional dynamics. In this section, we apply DMD to the simulated household panel $\{\mathbf{y}_t\}_{t=1}^T$ and evaluate how well it recovers the spectral, state, and cross-sectional properties of the underlying dynamics.

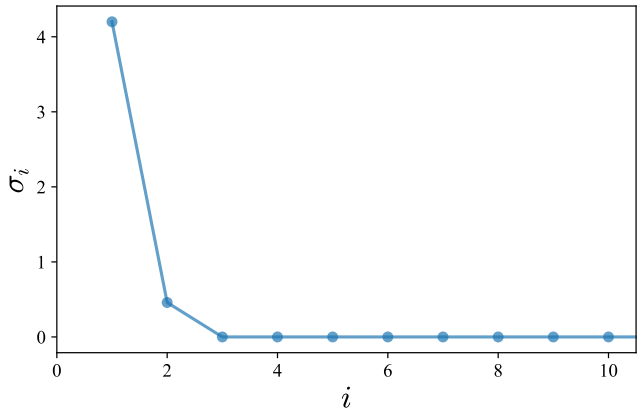


**Figure 5:** Scree plot for the household panel. The sharp elbow after the second singular value motivates $N = 2$.

We begin by selecting the DMD rank $N$. Figure 5 presents a scree plot of singular values of the centered snapshot matrix $\mathbf{Y}$ constructed from $\{\mathbf{y}_t\}_{t=1}^{T}$ by stacking the observations as columns. The sharp elbow after the second singular value suggests that $N = 2$ is a natural choice, consistent with the economic structure of the model (Section 3.3). We therefore set $N = 2$ in the DMD implementation below. With $N$ fixed, we first assess spectral recovery by comparing eigenvalues of the estimated DMD operator $\widehat{\mathbf{B}}$ to those of the true system. Figure 6 overlays the top 20 eigenvalues of $\mathbf{A}$, the eigenvalues of the population VAR(1) coefficient $\mathbf{B}^*$ in (43), the DMD eigenvalues, and the aggregate endowment persistence $\rho_1$. With $N = 2$, the DMD eigenvalues closely match those of $\mathbf{B}^*$, confirming that DMD recovers the population VAR(1) dynamics.

The theoretical connection between DMD modes and the centered measurement matrix $\mathbf{G}$ can be verified numerically. This example illustrates the rank-deficient measurement case described in Proposition 1 since the measurement matrix $\mathbf{G}$ has rank $2 < N_x$ as we discussed in Section 3.4. From (43), $\text{range}(\mathbf{B}^*) \subseteq \text{range}(\mathbf{G})$. In this experiment, the DMD modes $\mathbf{\Phi}$ are nearly collinear with the leading eigenvectors of $\mathbf{B}^*$. Computing the correlation between estimated DMD modes and the leading

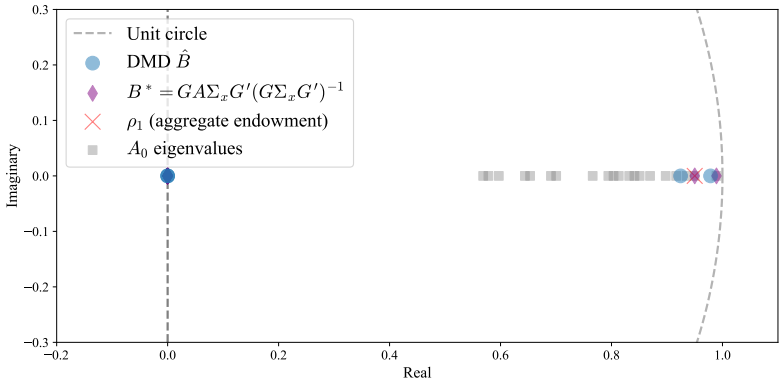


**Figure 6:** Spectral recovery

eigenvectors of $\mathbf{B}^*$, we find correlations of 0.97 for both modes, with principal angles between subspaces of less than $0.05°$. Projecting $\mathbf{\Phi}$ onto $\text{range}(\mathbf{G})$ yields a relative residual of about $10^{-12}$. As shown in Figure 4, the centered panel has only two singular values with magnitude larger than $1.0 \times 10^{-10}$, so DMD recovers exactly two modes corresponding to the most important aggregate state components.

Spectral recovery does not by itself imply that the associated low-dimensional DMD state tracks the latent state. As we discussed in Section 3.4, we expect the leading DMD modes to track the aggregate capital and endowment components of $\mathbf{x}_t$ if DMD recovers the low-dimensional dynamics from the household panel. Figure 7 examines whether the estimated DMD mode coordinates track the aggregate state. Let $\widetilde{\mathbf{x}}_t = \mathbf{\Phi}^+ \mathbf{y}_t$ denote the DMD mode coordinates. For each mode, we compute its correlation with every component of the true state $\mathbf{x}_t$ and plot it against the state component with the highest absolute correlation (both standardized as z-scores). Indeed, the first DMD mode closely tracks aggregate capital, and the second tracks aggregate endowment with very high correlation, confirming that DMD recovers the true low-dimensional dynamics embedded in the high-dimensional panel.

Beyond time-series recovery, we can ask whether the estimated modes reproduce the cross-sectional structure implied by the sharing rule. Figures 8 and 9 examine how

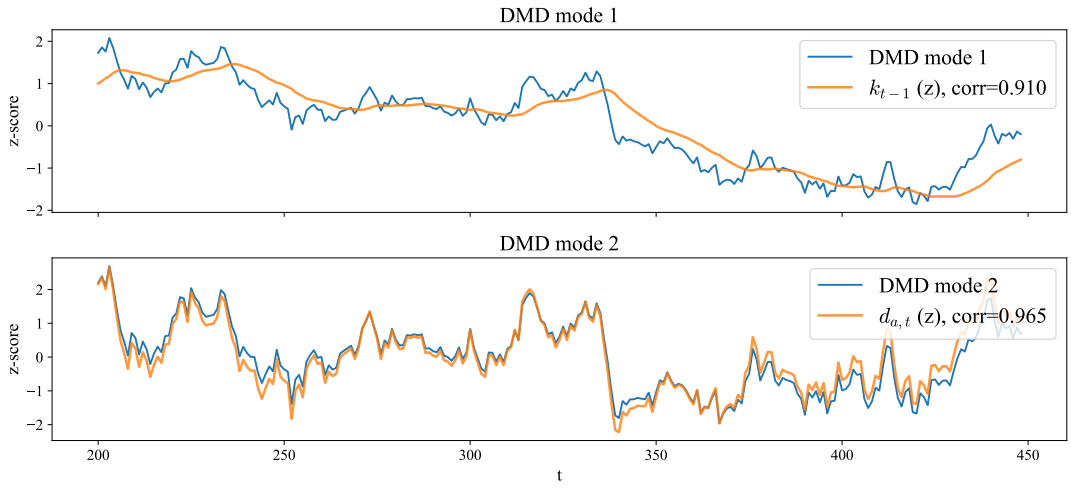


**Figure 7:** DMD mode time series for the stacked panel $\mathbf{y}_t$ compared with selected components of the true aggregate state $\mathbf{x}_t$.

DMD mode loadings vary across households. The DMD modes matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ has rows corresponding to observables. Under the sharing rule (24), the component of consumption driven by aggregate dynamics scales with $\mu_j$, so households with larger $\mu_j$ should load more heavily on aggregate modes. Figure 8 confirms this by plotting the magnitudes of DMD loadings for pre- and post-redistribution income and for consumption against household wealth percentiles. Denote these loadings by $\Phi_{jk}^{(y_{\text{pre}})}$, $\Phi_{jk}^{(y_{\text{post}})}$, and $\Phi_{jk}^{(c)}$, respectively, where households are ordered by the magnitude of their Gorman weights $\mu_j$. Figure 8 shows that households with higher percentiles have a larger loading on DMD modes of pre-redistribution income. Although the same trend holds for post-redistribution income and consumption, the relationship is less pronounced because redistribution attenuates inequality.

Figure 9 reinforces the idea that the loadings reflect the sharing rule. We extract the consumption block $\mathbf{\Phi}^{(c)}$ and plot $|\mathbf{\Phi}_{jk}^{(c)}|$ against household $j$'s Gorman weight $\mu_j$, showing that the correlation between $|\mathbf{\Phi}_{j1}^{(c)}|$ and $\mu_j$ is unity, indicating that DMD loadings recover the cross-sectional sharing rule.

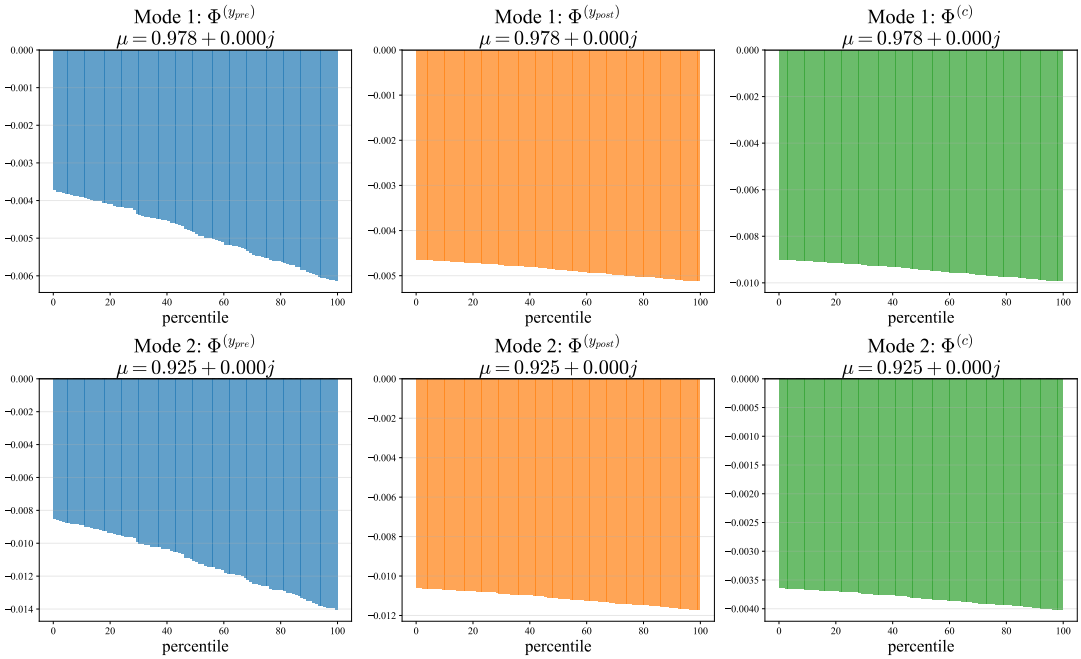


**Figure 8:** DMD loadings by wealth percentile for the stacked panel $\mathbf{y}_t$. Households are ordered by the Gorman weights $\{\mu_j\}$ (ascending).

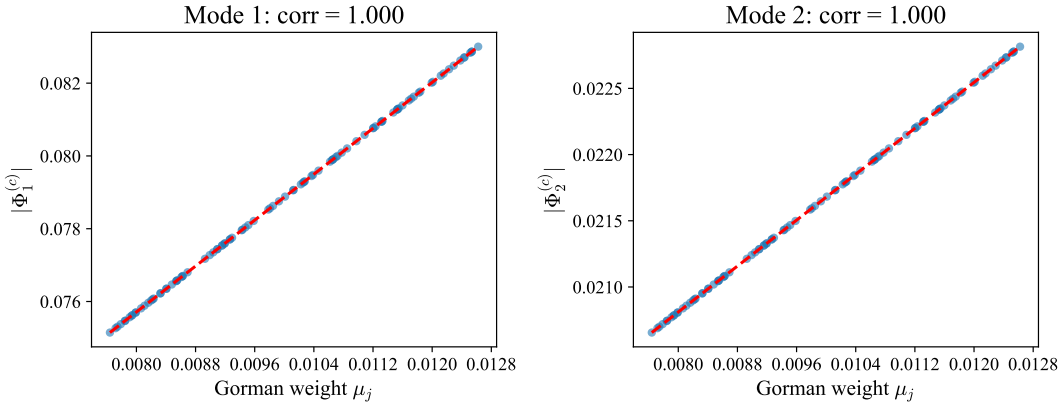


**Figure 9:** In the 100-household experiment, DMD loadings in the consumption block are proportional to the Gorman weights $\{\mu_j\}$, consistent with the sharing rule (24).

To move beyond in-sample diagnostics, Figure 10 evaluates forecasting performance using a train-test split with training fraction $0.7$. We estimate DMD on the training sample to obtain $\widehat{\mathbf{B}}$ and the training-sample mean $\bar{\mathbf{y}}_{\text{train}}$. For each test observation $\mathbf{y}_t^{\text{GL}}$, the $k$-step-ahead conditional-mean forecast is

$$\hat{\mathbf{y}}_{t+k|t}^{\text{GL}} = \bar{\mathbf{y}}_{\text{train}} + \widehat{\mathbf{B}}^k(\mathbf{y}_t^{\text{GL}} - \bar{\mathbf{y}}_{\text{train}}), \tag{48}$$

which iterates the estimated VAR operator $k$ steps forward. The figure plots the cross-sectional median of both realized values $\mathbf{y}_{t+k}^{\text{GL}}$ and forecasts $\hat{\mathbf{y}}_{t+k|t}^{\text{GL}}$ for horizons $k \in \{1, 3, 5\}$. Short-horizon predictions closely track the realized median dynamics, while longer-horizon predictions revert toward unconditional means as the stable DMD operator is iterated forward. The same pattern holds for all cross-sectional percentiles.

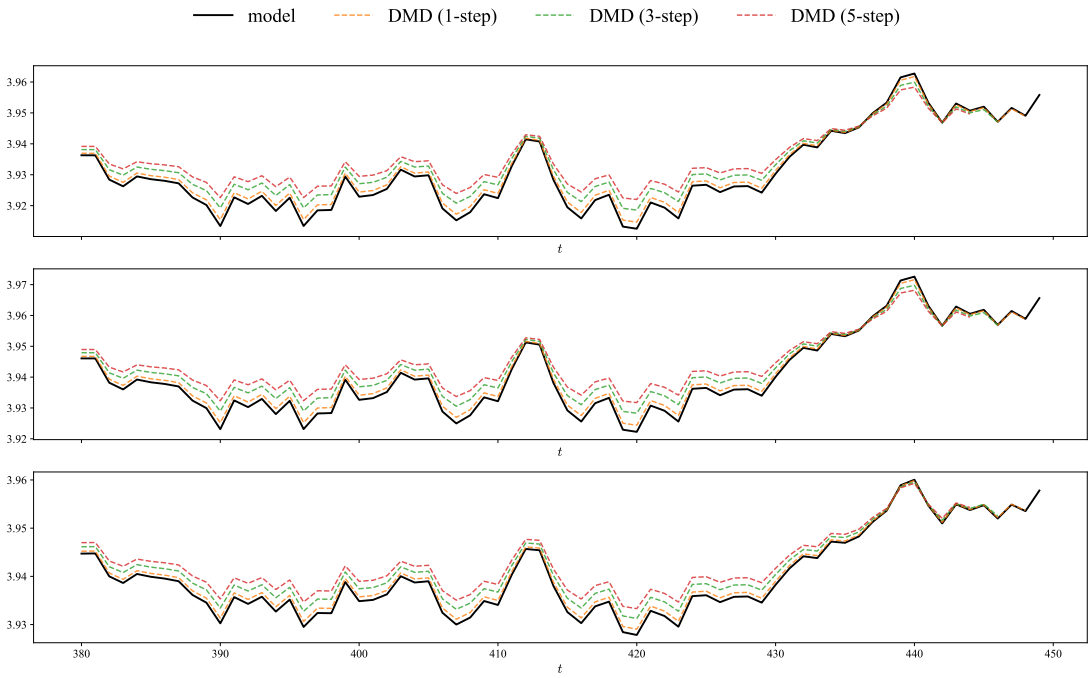


**Figure 10:** Out-of-sample DMD predictions in the 100-household experiment at horizons $k \in \{1, 3, 5\}$.

Finally, the Gorman laboratory provides an informative numerical example that

verifies key theoretical results in Section 2, particularly the implications of the rank-2 structure of the measurement matrix $\mathbf{G}$ documented in Figure 4. When $\text{rank}(\mathbf{G}) < N_x$, the population VAR coefficient can be written as $\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^+_{\mathbf{\Sigma}_x}$, where $\mathbf{G}^+_{\mathbf{\Sigma}_x}$ is a weighted pseudoinverse that depends on the state covariance $\mathbf{\Sigma}_x$. In this rank-deficient case, the standard Moore–Penrose pseudoinverse $\mathbf{G}^+$ is not guaranteed to recover $\mathbf{A}$ from $\mathbf{B}^*$, but a minimal two-dimensional state representation using $(k_{t-1}, d_{a,t})$ can be constructed to recover $\mathbf{A}$. Appendix B develops these results and verifies them numerically.

## 4 Other Applications of PCA

Stock and Watson (2016, Sec. 2) describe how a principal-components representation of the contemporaneous covariance matrix of a covariance-stationary vector stochastic process is an essential component of what they call a "static" representation of a "dynamic" factor model. That application of principal components differs substantially from ours. To appreciate this, notice where singular value decompositions make two appearances in our calculations.[15] We use a reduced SVD $\mathbf{Y} = \widetilde{\mathbf{U}}\,\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^\top$ to form submatrices $\mathbf{U}$ of $\widetilde{\mathbf{U}}$ and $\mathbf{V}$ of $\widetilde{\mathbf{V}}$ that appear in (13). This yields the rank-$N$ truncation $\mathbf{Y}_N := \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^\top$, which we then use to form $\widehat{\mathbf{B}} = \mathbf{Y}'\,\mathbf{Y}_N^+$. We use $\mathbf{U}$ and $\mathbf{V}$ again when we form the reduced operator $\widetilde{\mathbf{A}} = \mathbf{U}^\top\,\widehat{\mathbf{B}}\,\mathbf{U} \in \mathbb{R}^{N\times N}$. At this point we compute an eigendecomposition

$$\widetilde{\mathbf{A}} = \mathbf{W}\,\mathbf{\Lambda}\,\mathbf{W}^{-1},$$

where columns of the $N \times N$ matrix $\mathbf{W}$ are eigenvectors of $\widetilde{\mathbf{A}}$ and eigenvalues of $\widetilde{\mathbf{A}}$ appear on the diagonal of the diagonal matrix $\mathbf{\Lambda}$.

By way of contrast, Stock and Watson describe how singular value decompositions are applied to construct a "static" representation of a "dynamic" factor model by (i)

[15] Many modern procedures compute eigendecompositions by first computing a singular value decomposition. For example, see Strang (2020).

assuming that an observed process $\{y_t\}$ is covariance-stationary, (ii) estimating the $M \times M$ covariance matrix $C_{\mathbf{y}}$ of $y_t$, (iii) using a singular value decomposition to compute $M$ principal components of $C_{\mathbf{y}}$, then selecting the $r$ most important components, then (iv) adopting statistical detection procedures to infer $q < r$ "factors" in what Stock and Watson call an associated "dynamic" factor model. In Sections 2–2.3, the object targeted by DMD is the lag-1 linear projection coefficient $\mathbf{B}^*$. The rank-$N$ DMD operator is a low-rank estimator of this projector computed from snapshots.

A reader of Stock and Watson (2016) will recognize how our procedure instead resides within a distinct tradition that connects LQG hidden Markov models and vector autoregressions to infer parameters of the hidden Markov model, a tradition in which a principal components analysis rarely makes an appearance. See especially Stock and Watson (2016, Sec. 2). As Section 2.3 above indicates, the hidden Markov model affiliated with our DMD procedure is a special case of the state-space models described in Stock and Watson (2016, Sec. 2).

PCA estimators of factors and the associated loadings are identified only up to a scale and rotation. Bai and Ng (2013) propose three sets of restrictions that imply exact identification of the factors and loadings resulting from the PC estimator.

The first set of restrictions for PC requires that both the factors and loadings are orthogonal. The second requires that the factors are orthogonal and that the top $N \times N$ block of $\mathbf{G}$ is lower triangular. The third leaves the factors unrestricted but requires the top $N \times N$ block of $\mathbf{G}$ to be the identity matrix. In sum, identification for principal components requires that either the factors are orthogonal, the loadings are orthogonal, or the loadings have special structure implying that one or more observations are noisy measurements of the factors.

DMD yields a modal state representation of the fitted reduced-rank VAR(1) on the DMD subspace, with diagonal $\mathbf{\Lambda}$. The associated residuals are pseudo-innovations unless the data are generated from VAR(1) with innovations orthogonal to the full

past (as in the injective-measurement case of Proposition 2). Accordingly, the nonzero eigenvalues in $\mathbf{\Lambda}$ estimate eigenvalues of the lag-1 linear projection coefficient $\mathbf{B}^*$. They coincide with eigenvalues of the structural transition matrix $\mathbf{A}$ only under full-column-rank $\mathbf{G}$ (Proposition 2(ii)). Our framework does not require that the columns of $\mathbf{G}$ are orthogonal, nor does it require that one or more observations are noisy measurements of the factors. We summarize these comparisons in Table 1.

| **Identifying restrictions** | | |
|---|---|---|
| | **Principal components** | **DMD** |
| PC Option 1 | • $\frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{x}}_t\widetilde{\mathbf{x}}_t^\top = I_N$<br>• $\mathbf{G}^\top\mathbf{G}$ is a diagonal matrix with distinct entries | • Fitted reduced-rank VAR(1) on DMD subspace with<br>– $\mathbf{\Lambda}$ diagonal (in modal coordinates)<br>– $\mathbf{C}$ unrestricted<br>• Eigenvalues of $\mathbf{\Lambda}$ estimate $\mathbf{B}^*$; equal $\mathbf{A}$ eigenvalues under full-column-rank $\mathbf{G}$ |
| PC Option 2 | • $\frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{x}}_t\widetilde{\mathbf{x}}_t^\top = I_N$<br>• $\mathbf{G} = \begin{pmatrix}\mathbf{G}_1\\ \mathbf{G}_2\end{pmatrix}$<br>• $\mathbf{G}_1 = \begin{pmatrix}\mathbf{g}_{11} & 0 & \cdots & 0\\ \mathbf{g}_{21} & \mathbf{g}_{22} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{g}_{r1} & \mathbf{g}_{r2} & \cdots & \mathbf{g}_{rr}\end{pmatrix}, \mathbf{g}_{ii} \neq 0, i = 1,\ldots,r$ | |
| PC Option 3 | • $\widetilde{\mathbf{x}}_t$ unrestricted<br>• $\mathbf{G} = \begin{pmatrix}I_N\\ \mathbf{G}_2\end{pmatrix}$ | |

**Table 1:** Identifying restrictions for principal components and DMD

## 5 Concluding Remarks

This paper connects Dynamic Mode Decomposition to reduced-rank VARs and linear state-space models, providing both theoretical foundations and empirical validation on high-dimensional household panel data generated by a heterogeneous-agent economy.

Section 2 establishes the population theory. Proposition 1 shows that the lag-1 linear

projection coefficient $\mathbf{B}^*$ satisfies $\text{rank}(\mathbf{B}^*) \leq \text{rank}(\mathbf{G})$, so that $\mathbf{B}^*$ captures dynamics on the support subspace of observables. Proposition 2 establishes conditions under which the nonzero eigenvalues of $\mathbf{B}^*$ coincide with those of the latent transition matrix $\mathbf{A}$: when $\mathbf{G}$ has full column rank and $\mathbf{\Sigma}_x \succ 0$, we have $\mathbf{A} = \mathbf{G}^+ \mathbf{B}^* \mathbf{G}$. Propositions 3 and 4 characterize DMD as a rank-$N$ estimator of $\mathbf{B}^*$ and interpret the fitted residuals as pseudo-innovations.

Section 3 applies DMD to a heterogeneous-agent economy with Gorman aggregation, where the true aggregate dynamics are known. Using only household-level income and consumption for 100 households, DMD with two modes recovers eigenvalues that closely match those of $\mathbf{B}^*$ and align with the dominant eigenvalues of $\mathbf{A}$. The estimated mode time series correlate 0.91 and 0.96 with aggregate capital and the endowment state, respectively, while loadings correlate with the Gorman weights that govern redistribution perfectly. Section 3.4 provides an information-theoretic interpretation: $\mathbf{B}^*$ is the KL projection of the one-step-ahead conditional law onto a Gaussian VAR(1) family.

Section 4 clarifies how DMD differs from principal components analysis in dynamic factor models. While PCA targets static covariance structure, DMD exploits temporal dependence to identify dynamic factors via a reduced-rank VAR, yielding different identification restrictions.

The analysis also highlights limitations. When measurement is rank-deficient, DMD recovers the dynamics visible in observables but cannot identify the full latent transition matrix without additional restrictions. Nevertheless, DMD provides a computationally efficient approach to extracting low-dimensional dynamics from high-dimensional economic panels, and we expect it to prove useful in applications ranging from household surveys to firm-level data.

# A Appendix

## A.1 Proofs

### A.1.1 Proposition 1

*Proof.* For (i), since $\mathbf{y}_t = \mathbf{G}\,\mathbf{x}_t$ and $\mathbf{x}_t$ is stationary,

$$\mathrm{Cov}(\mathbf{y}_t, \mathbf{y}_{t-1}) = \mathbf{G}\,\mathbf{A}\,\mathbf{\Sigma}_x\,\mathbf{G}^\top, \qquad \mathrm{Cov}(\mathbf{y}_{t-1}) = \mathbf{G}\,\mathbf{\Sigma}_x\,\mathbf{G}^\top.$$

Substituting into (5) yields $\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\mathbf{\Sigma}_x\,\mathbf{G}^\top(\mathbf{G}\,\mathbf{\Sigma}_x\,\mathbf{G}^\top)^+ = \mathbf{G}\,\mathbf{A}\,\mathbf{G}_{\mathbf{\Sigma}_x}^+$.

(ii) follows from $\mathbf{B}^* = \mathbf{G}(\mathbf{A}\,\mathbf{G}_{\mathbf{\Sigma}_x}^+)$, and if $\mathbf{B}^*\,\mathbf{v} = \lambda\,\mathbf{v}$ with $\lambda \neq 0$ then $\mathbf{v} = (1/\lambda)\,\mathbf{B}^*\,\mathbf{v} \in \mathrm{range}(\mathbf{G})$. □

### A.1.2 Proposition 2

*Proof.* Regarding (i), since $\mathbf{G}$ has full column rank, $\mathbf{G}^+\,\mathbf{G} = \mathbf{I}_{N_x}$ and therefore $\mathbf{x}_{t-1} = \mathbf{G}^+\,\mathbf{y}_{t-1}$. Using $\mathbf{x}_t = \mathbf{A}\,\mathbf{x}_{t-1} + \mathbf{C}\,\mathbf{w}_t$ gives

$$\mathbf{y}_t = \mathbf{G}\,\mathbf{x}_t = \mathbf{G}(\mathbf{A}\,\mathbf{x}_{t-1} + \mathbf{C}\,\mathbf{w}_t) = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^+\,\mathbf{y}_{t-1} + \mathbf{G}\,\mathbf{C}\,\mathbf{w}_t = \widetilde{\mathbf{B}}\,\mathbf{y}_{t-1} + \mathbf{u}_t\,.$$

Because $\{\mathbf{w}_t\}$ is i.i.d. and independent of the past $\mathbf{y}^{t-1}$, the same holds for $\{\mathbf{u}_t\}$, so $\mathbf{u}_t \perp \mathbf{y}^{t-1}$. Hence $\mathbb{E}[\mathbf{y}_t \mid \mathbf{y}^{t-1}] = \widetilde{\mathbf{B}}\,\mathbf{y}_{t-1}$ and $\mathbf{a}_t = \mathbf{y}_t - \mathbb{E}[\mathbf{y}_t \mid \mathbf{y}^{t-1}] = \mathbf{u}_t$. This completes the first part of (i).

Let $\mathbf{\Gamma}_0 := \mathrm{Cov}(\mathbf{y}_{t-1})$ and $\mathbf{\Gamma}_1 := \mathrm{Cov}(\mathbf{y}_t, \mathbf{y}_{t-1})$. Orthogonality implies $\widetilde{\mathbf{B}}\mathbf{\Gamma}_0 = \mathbf{\Gamma}_1$.[16] By definition $\mathbf{B}^* = \mathbf{\Gamma}_1\mathbf{\Gamma}_0^+$, so $(\mathbf{B}^* - \widetilde{\mathbf{B}})\mathbf{\Gamma}_0 = 0$. Since $\mathbf{y}_{t-1} \in \mathrm{range}(\mathbf{\Gamma}_0)$ a.s., we have $(\mathbf{B}^* - \widetilde{\mathbf{B}})\,\mathbf{y}_{t-1} = 0$ a.s. and hence statement (i) follows.

For (ii), by Proposition 1(i), $\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\mathbf{\Sigma}_x\,\mathbf{G}^\top(\mathbf{G}\,\mathbf{\Sigma}_x\,\mathbf{G}^\top)^+$. If $\mathbf{\Sigma}_x \succ 0$ and $\mathbf{G}$ has

[16] This can be seen from the fact that $\mathbf{\Gamma}_1 = \mathbb{E}[\mathbf{y}_t\,\mathbf{y}_{t-1}^\top] = \mathbb{E}[(\widetilde{\mathbf{B}}\,\mathbf{y}_{t-1} + \mathbf{u}_t)\,\mathbf{y}_{t-1}^\top] = \widetilde{\mathbf{B}}\,\mathbb{E}[\mathbf{y}_{t-1}\,\mathbf{y}_{t-1}^\top] + \mathbb{E}[\mathbf{u}_t\,\mathbf{y}_{t-1}^\top] = \widetilde{\mathbf{B}}\mathbf{\Gamma}_0 + 0$.

full column rank, then

$$(\mathbf{G}\,\boldsymbol{\Sigma}_x\,\mathbf{G}^\top)^+ = (\mathbf{G}^+)^\top \boldsymbol{\Sigma}_x^{-1}\,\mathbf{G}^+,$$

so $\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\boldsymbol{\Sigma}_x\,\mathbf{G}^\top(\mathbf{G}^+)^\top\boldsymbol{\Sigma}_x^{-1}\,\mathbf{G}^+ = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^+ = \widetilde{\mathbf{B}}$.[17] Then $\mathbf{A} = \mathbf{G}^+\,\mathbf{B}^*\,\mathbf{G}$ follows from $\mathbf{G}^+\,\mathbf{G} = \mathbf{I}_{N_x}$. □

### A.1.3 Corollary 1

*Proof.* Fix any symmetric $\boldsymbol{\Sigma} \succeq 0$ and let $\boldsymbol{\Sigma}^{1/2}$ denote its symmetric square root. Define $\mathbf{Q} := \mathbf{G}\,\boldsymbol{\Sigma}^{1/2}$. Since $\mathbf{G}$ has full column rank, $\mathbf{G}^\top$ is surjective, and therefore

$$\mathrm{range}(\mathbf{Q}^\top) = \mathrm{range}(\boldsymbol{\Sigma}^{1/2}\,\mathbf{G}^\top) = \mathrm{range}(\boldsymbol{\Sigma}^{1/2}).$$

Hence $\mathbf{Q}^\top(\mathbf{Q}\,\mathbf{Q}^\top)^+\,\mathbf{Q}$ is the orthogonal projector onto $\mathrm{range}(\boldsymbol{\Sigma}^{1/2})$, so it acts as the identity on $\mathrm{range}(\boldsymbol{\Sigma}^{1/2})$. Hence,

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\;\mathbf{Q}^\top(\mathbf{Q}\,\mathbf{Q}^\top)^+\,\mathbf{Q}\;\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}\;\mathbf{G}^\top(\mathbf{G}\,\boldsymbol{\Sigma}\,\mathbf{G}^\top)^+\;\mathbf{G}\;\boldsymbol{\Sigma}.$$

Applying this identity with $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_\infty$ in (4) yields

$$\boldsymbol{\Sigma}_\infty = \mathbf{A}\,\boldsymbol{\Sigma}_\infty\,\mathbf{A}^\top + \mathbf{C}\,\mathbf{C}^\top - \mathbf{A}\,\boldsymbol{\Sigma}_\infty\,\mathbf{G}^\top(\mathbf{G}\,\boldsymbol{\Sigma}_\infty\,\mathbf{G}^\top)^+\,\mathbf{G}\,\boldsymbol{\Sigma}_\infty\,\mathbf{A}^\top = \mathbf{C}\,\mathbf{C}^\top,$$

which proves (8). Under the additional assumption $\mathbf{C}\,\mathbf{C}^\top \succ 0$, we have $\boldsymbol{\Sigma}_\infty \succ 0$. Letting $\mathbf{Q} := \mathbf{G}\,\boldsymbol{\Sigma}_\infty^{1/2}$, $\mathbf{Q}$ has full column rank, so $\mathbf{Q}^\top(\mathbf{Q}\,\mathbf{Q}^\top)^+\,\mathbf{Q} = \mathbf{I}_{N_x}$, which is equivalent to

$$\boldsymbol{\Sigma}_\infty\,\mathbf{G}^\top(\mathbf{G}\,\boldsymbol{\Sigma}_\infty\,\mathbf{G}^\top)^+\,\mathbf{G} = \mathbf{I}_{N_x}.$$

Multiplying (3) on the right by $\mathbf{G}$ gives $\mathbf{K}\,\mathbf{G} = \mathbf{A}$. □

[17] Because $\mathbf{G}^\top(\mathbf{G}^+)^\top = (\mathbf{G}^+\,\mathbf{G})^\top = \mathbf{I}_{N_x}$.

### A.1.4 Proposition 3

*Proof.* Recall that $\mathbf{Y}_N := \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^\top$ denotes the rank-$N$ truncated SVD of the snapshot matrix. Assume $\sigma_N > 0$. Then

$$\mathbf{Y}_N^+ = \mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{U}^\top, \qquad \widehat{\mathbf{B}} = \mathbf{Y}'\,\mathbf{Y}_N^+ = \mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{U}^\top.$$

Define the reduced operator and the DMD modes by

$$\widetilde{\mathbf{A}} := \mathbf{U}^\top\,\mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}, \quad \widetilde{\mathbf{A}}\,\mathbf{W} = \mathbf{W}\,\mathbf{\Lambda}, \quad \mathbf{\Phi} := \mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{W}.$$

Then

$$\widehat{\mathbf{B}}\mathbf{\Phi} = \mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}(\mathbf{U}^\top\,\mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1})\,\mathbf{W} = \mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}(\widetilde{\mathbf{A}}\,\mathbf{W}) = \mathbf{Y}'\,\mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{W}\,\mathbf{\Lambda} = \mathbf{\Phi}\mathbf{\Lambda}.$$

This proves statement (i).

When restricting $\widehat{\mathbf{B}}$ to act on $\mathrm{col}(\mathbf{\Phi})$, equivalently replace $\widehat{\mathbf{B}}$ by $\widehat{\mathbf{B}}\mathbf{P}_{\mathbf{\Phi}}$. Using $\mathbf{Y}_N^+\,\mathbf{\Phi} = \mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{U}^\top\,\mathbf{\Phi} = \mathbf{V}\,\mathbf{\Sigma}^{-1}\widetilde{\mathbf{A}}\,\mathbf{W} = \mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{W}\,\mathbf{\Lambda}$, we have

$$\widehat{\mathbf{B}}\mathbf{P}_{\mathbf{\Phi}} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^+,$$

and hence $\widehat{\mathbf{B}}$ and $\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^+$ coincide on $\mathrm{col}(\mathbf{\Phi})$.[18] This completes the proof. □

# B Minimal State Recovery in Gorman Application

Recall that with $\mathbf{\Sigma}_x = \mathrm{Cov}(\mathbf{x}_t)$, the population VAR(1) coefficient in (43) can be written as

$$\mathbf{B}^* = \mathbf{G}\,\mathbf{A}\,\mathbf{G}_{\mathbf{\Sigma}_x}^+, \quad \text{where} \quad \mathbf{G}_{\mathbf{\Sigma}_x}^+ := \mathbf{\Sigma}_x\,\mathbf{G}^\top\big(\mathbf{G}\,\mathbf{\Sigma}_x\,\mathbf{G}^\top\big)^+ \tag{49}$$

[18] Moreover, among all $\mathbf{B}$ with $\mathbf{B}\,\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda}$, the unique minimizer of the Frobenius norm $\|\,\mathbf{B}\,\|_F$ is $\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^+$. This follows from the KKT conditions for $\min_{\mathbf{B}} \frac{1}{2}\|\,\mathbf{B}\,\|_F^2$ subject to $\mathbf{B}\,\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda}$.

is the $\mathbf{\Sigma}_x$-weighted pseudoinverse of $\mathbf{G}$. Equation (49) is simply a re-expression of Proposition 1(i).

When $\mathbf{G}$ has full column rank and $\mathbf{\Sigma}_x$ is positive definite, Proposition 2(ii) implies that $\mathbf{G}^{+}_{\mathbf{\Sigma}_x} = \mathbf{G}^{+}$ and hence $\mathbf{B}^{*} = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^{+}$. When $\mathrm{rank}(\mathbf{G}) < N_x$, (1) is partially observed and $\mathbf{G}^{+}_{\mathbf{\Sigma}_x}$ need not equal $\mathbf{G}^{+}$; in general, the simplification $\mathbf{B}^{*} = \mathbf{G}\,\mathbf{A}\,\mathbf{G}^{+}$ fails and $\mathbf{A}$ cannot be recovered from $(\mathbf{B}^{*}, \mathbf{G})$ via $\mathbf{A} = \mathbf{G}^{+}\,\mathbf{B}^{*}\,\mathbf{G}$.

We verify these theoretical results numerically in the Gorman experiment. Table 2 compares $\mathbf{B}^{*}$ to the projections $\mathbf{G}\,\mathbf{A}\,\mathbf{G}^{+}$ and $\mathbf{G}\,\mathbf{A}\,\mathbf{G}^{+}_{\mathbf{\Sigma}_x}$. In the full system with the high-dimensional aggregate state, the centered state has dimension $N_x = 154$ and the measurement matrix satisfies $\mathrm{rank}(\mathbf{G}) = 2 < N_x$, reflecting the rank-2 structure discussed in Section 3.3 (see Figure 4). Consistent with Proposition 1, $\mathbf{B}^{*}$ is matched by $\mathbf{G}\,\mathbf{A}\,\mathbf{G}^{+}_{\mathbf{\Sigma}_x}$ up to numerical precision, while $\mathbf{G}\,\mathbf{A}\,\mathbf{G}^{+}$ can deviate sharply when $\mathbf{G}$ is rank-deficient.

**Table 2:** Comparison of population equalities and estimates in the Gorman experiment

| | **Minimal state** $(k_{t-1}, d_{a,t})$ | **Full state** (154-dimensional) |
|---|---|---|
| State dimension $N_x$ | 2 | 154 |
| Numerical rank($\mathbf{G}$) | 2 (full) | 2 (deficient) |
| $\|\,\mathbf{B}^{*} - \mathbf{G}\,\mathbf{A}\,\mathbf{G}^{+}\,\|_F / \|\,\mathbf{B}^{*}\,\|_F$ | $2.8 \times 10^{-14}$ | 31.4 |
| $\|\,\mathbf{B}^{*} - \mathbf{G}\,\mathbf{A}\,\mathbf{G}^{+}_{\mathbf{\Sigma}_x}\,\|_F / \|\,\mathbf{B}^{*}\,\|_F$ | $1.8 \times 10^{-14}$ | $1.8 \times 10^{-14}$ |
| $\|\,\mathbf{G}^{+}\,\mathbf{B}^{*}\,\mathbf{G} - \mathbf{A}\,\|_F / \|\,\mathbf{A}\,\|_F$ | $1.9 \times 10^{-14}$ | 0.982 |
| $\|\,\mathbf{G}^{+}\,\widehat{\mathbf{B}}\,\mathbf{G} - \mathbf{A}\,\|_F / \|\,\mathbf{A}\,\|_F$ | 0.0294 | 0.989 |

To illustrate the full-column-rank case, we also construct a reduced ("minimal") state representation using only the two aggregate factors that drive the centered panel, namely, aggregate capital $k_{t-1}$ and the aggregate endowment state $d_{a,t}$. Let $\mathbf{x}^{\min}_t := (k_{t-1}, d_{a,t})^{\top}$ and write the centered panel as $\mathbf{y}_t = \mathbf{G}_{\min}\,\mathbf{x}^{\min}_t$, where $\mathbf{G}_{\min} \in \mathbb{R}^{M \times 2}$ has full column rank. In our experiment, $(M, T) = (300, 250)$, so $\mathbf{G}_{\min}$ has shape $300 \times 2$ and the full data matrix $[\mathbf{y}_1, \ldots, \mathbf{y}_{T+1}]$ has shape $300 \times 251$. Since $\mathbf{G}_{\min}$ has full column rank, Proposition 1 implies $\mathbf{B}^{*} = \mathbf{G}_{\min}\,\mathbf{A}_{\min}\,\mathbf{G}^{+}_{\min}$ and $\mathbf{A}_{\min} = \mathbf{G}^{+}_{\min}\,\mathbf{B}^{*}\,\mathbf{G}_{\min}$.

The matrix $\mathbf{G}_{\text{min}}\,\widehat{\mathbf{\Sigma}}_{x,\text{min}}\,\mathbf{G}_{\text{min}}^{\top}$ has two singular values above numerical precision (approximately $1.19 \times 10^{-1}$ and $1.26 \times 10^{-3}$), confirming that the minimal-state observables are effectively two-dimensional. Likewise, the demeaned snapshot matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_T]$ has only two singular values above $10^{-10}$ (approximately $14.63$ and $1.51$), with all remaining singular values at machine precision. Applying DMD to $\{\mathbf{y}_t\}$ in this minimal-state representation yields $\|\widehat{\mathbf{B}} - \mathbf{B}^*\|_F / \|\mathbf{B}^*\|_F = 4.80 \times 10^{-3}$. The recovered transition matrix is accurate up to machine precision with $\|\mathbf{G}_{\text{min}}^{+}\,\mathbf{B}^*\,\mathbf{G}_{\text{min}} - \mathbf{A}_{\text{min}}\|_F / \|\mathbf{A}_{\text{min}}\|_F = 1.9 \times 10^{-14}$. By contrast, a naive attempt to recover the full high-dimensional transition matrix via $\mathbf{A} = \mathbf{G}^{+}\,\mathbf{B}^*\,\mathbf{G}$ produces large errors in this experiment $\|\mathbf{G}^{+}\,\mathbf{B}^*\,\mathbf{G} - \mathbf{A}\|_F / \|\mathbf{A}\|_F \approx 0.982$. Similarly, using the DMD estimator $\widehat{\mathbf{B}}$ instead of $\mathbf{B}^*$, the minimal transition matrix estimate is given by

$$\widehat{\mathbf{A}}_{\text{min}} := \mathbf{G}_{\text{min}}^{+}\,\widehat{\mathbf{B}}\,\mathbf{G}_{\text{min}} = \begin{bmatrix} 0.989 & 0.392 \\ -0.002 & 0.914 \end{bmatrix}$$

is close to the true minimal transition matrix

$$\mathbf{A}_{\text{min}} := \begin{bmatrix} 0.989 & 0.392 \\ 0.000 & 0.950 \end{bmatrix}$$

with relative error $\|\mathbf{G}_{\text{min}}^{+}\,\widehat{\mathbf{B}}\,\mathbf{G}_{\text{min}} - \mathbf{A}_{\text{min}}\|_F / \|\mathbf{A}_{\text{min}}\|_F = 0.0294$, which is lower than the relative error $0.989$ for the full state. Table 2 summarizes these findings in this particular application.

## B.1 Derivations for Section 2.1

### B.1.1 Derivation of (15)

Iterate equation (14) forward:

$$\widetilde{\mathbf{x}}_{t+j} = \mathbf{\Lambda}\widetilde{\mathbf{x}}_{t+j-1} + \mathbf{\Phi}^{+}\widehat{\mathbf{a}}_{t+j} \tag{50}$$

$$\widetilde{\mathbf{x}}_{t+j} = \mathbf{\Lambda}^{2}\widetilde{\mathbf{x}}_{t+j-2} + \mathbf{\Lambda}\mathbf{\Phi}^{+}\widehat{\mathbf{a}}_{t+j-1} + \mathbf{\Phi}^{+}\widehat{\mathbf{a}}_{t+j} \tag{51}$$

$$= \mathbf{\Lambda}^{j}\widetilde{\mathbf{x}}_{t} + \sum_{s=0}^{j-1}\mathbf{\Lambda}^{s}\mathbf{\Phi}^{+}\widehat{\mathbf{a}}_{t+j-s}. \tag{52}$$

### B.1.2 Derivation of (16)

The conditional expectation given $\widetilde{\mathbf{x}}_t$ is $\mathbb{E}[\widetilde{\mathbf{x}}_{t+j} \mid \widetilde{\mathbf{x}}_t] = \mathbf{\Lambda}^{j}\widetilde{\mathbf{x}}_t$. Thus the conditional covariance is

$$\mathbb{E}\left[(\widetilde{\mathbf{x}}_{t+j} - \mathbb{E}[\widetilde{\mathbf{x}}_{t+j} \mid \widetilde{\mathbf{x}}_t])(\widetilde{\mathbf{x}}_{t+j} - \mathbb{E}[\widetilde{\mathbf{x}}_{t+j} \mid \widetilde{\mathbf{x}}_t])^{\top}\right] \tag{53}$$

$$= \mathbb{E}\left[\Big(\sum_{s=0}^{j-1}\mathbf{\Lambda}^{s}\mathbf{\Phi}^{+}\widehat{\mathbf{a}}_{t+j-s}\Big)\Big(\sum_{r=0}^{j-1}\mathbf{\Lambda}^{r}\mathbf{\Phi}^{+}\widehat{\mathbf{a}}_{t+j-r}\Big)^{\top}\right] \tag{54}$$

$$= \sum_{s=0}^{j-1}\mathbf{\Lambda}^{s}\mathbf{\Phi}^{+}\widehat{\mathbf{\Omega}}(\mathbf{\Phi}^{+})^{\top}(\mathbf{\Lambda}^{\top})^{s}. \tag{55}$$

# C Parameter Values for the 100-Household Example

Table 3 lists the parameter values used in the 100-household Gorman economy of Section 3.3. Notation follows the main text. In particular, $\Lambda_s$, $\Pi_s$, $\Delta_h$, and $\Theta_h$ appear in the service equations following (20), while $\Delta_k$, $\Theta_k$, and $\Gamma$ appear in the planner's constraints following (22).

**Table 3:** Parameter values for the 100-household example

| | | |
|---|---|---|
| | *Preferences and technology (scalar)* | |
| $\beta$ | Discount factor | $0.952$ |
| $\Lambda_s$ | Service-habit loading | $0$ |
| $\Pi_s$ | Service-consumption loading | $1$ |
| $\Delta_h$ | Durable persistence | $0$ |
| $\Theta_h$ | Durable-consumption loading | $0$ |
| $\Delta_k$ | Capital persistence | $0.95$ |
| $\Theta_k$ | Capital-investment loading | $1$ |
| $\Gamma$ | Capital productivity | $0.1$ |
| | *Aggregate endowment process* | |
| $\rho_1$ | AR(1) coefficient | $0.95$ |
| $\sigma_a$ | Aggregate shock std. dev. | $0.5$ |
| | *Household heterogeneity* | |
| $J$ | Number of households | $100$ |
| $\alpha_j$ | Mean endowment | $\sim \mathcal{U}[3, 5]$ |
| $\phi_j$ | Aggregate exposure | $\sim \mathcal{U}[0.5, 1.5]$, $\sum_j \phi_j = 1$ |
| | *Idiosyncratic endowment shocks* | |
| $J_a$ | Absorbing households | $50$ |
| $\sigma_j$ | Idio. shock std. dev. | $0.2 + 4.8\, p_j^2$ |
| $\rho_j^d$ | Idio. shock persistence | $0.9\, p_j$ |

*Notes:* $p_j := 1 -$ percentile rank of $\alpha_j$. Preference shocks are muted (zero loadings). See Section 3.3 for details.

## References


BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

BAI, J. AND S. NG (2013): "Principal components estimation and identification of static factors," *Journal of Econometrics*, 176, 18–29.

BRUNTON, S. L., B. W. BRUNTON, J. L. PROCTOR, AND J. N. KUTZ (2016): "Koopman Invariant Subspaces and Finite Linear Representations of Nonlinear Dynamical Systems for Control," *PLOS ONE*, 11, e0150171.

BRUNTON, S. L. AND J. N. KUTZ (2022): *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, Cambridge University Press.

BURNS, A. F. AND W. C. MITCHELL (1946): *Measuring Business Cycles*, no. 2 in Studies in Business Cycles, National Bureau of Economic Research.

CSISZÁR, I. AND F. MATUS (2003): "Information projections revisited," *IEEE Transactions on Information Theory*, 49, 1474–1490.

EBERLY, J. AND N. WANG (2025): "Reallocating and Pricing Illiquid Capital: Two Productive Trees," *International Economic Review*, 66, 1887–1907.

GEWEKE, J. (1977): "The dynamic factor analysis of economic time series," in *Latent Variables in Socio-Economic Models*, ed. by D. Aigner and A. S. Goldberger, North-Holland.

GEWEKE, J. F. AND K. J. SINGLETON (1981): "Maximum likelihood "confirmatory" factor analysis of economic time series," *International Economic Review*, 37–54.

GOLUB, G. H. AND C. F. VAN LOAN (2013): *Matrix Computations*, Baltimore, MD: Johns Hopkins University Press, 4 ed.

GORMAN, W. M. (1953): "Community preference fields," *Econometrica: Journal of the Econometric Society*, 21, 63–80.

HANSEN, L. P. AND T. J. SARGENT (1993): "Seasonality and Approximation Errors in Rational Expectations Models," *Journal of Econometrics*, 55, 21–55.

——— (2013): *Recursive Models of Dynamic Linear Economies*, Princeton, NJ: Princeton University Press.

HIRSH, S. M., K. D. HARRIS, J. N. KUTZ, AND B. W. BRUNTON (2020): "Centering Data Improves the Dynamic Mode Decomposition," *SIAM Journal on Applied Dynamical Systems*, 19, 1920–1955.

HOOD, W. C. AND T. C. KOOPMANS (1953): *Studies in econometric method*, Wiley.

KOOPMANS, T. C. (1947): "Measurement without theory," *The Review of Economics and Statistics*, 29, 161–172.

——— (1950): *Statistical inference in dynamic economic models*, New York: Wiley.

LUCAS, JR., R. E. (1978): "Asset prices in an exchange economy," *Econometrica: Journal of the Econometric Society*, 46, 1429–1445.

LUCAS, R. E. J. (1987): *Models of business cycles*, vol. 26, Oxford Blackwell.

MARSCHAK, J. (1953): "Economic Measurement for Policy and Prediction," in *Studies in econometric method*, ed. by W. C. Hood and T. C. Koopmans, Wiley, 1–26.

MEZIC, I. (2020): "On Numerical Approximations of the Koopman Operator," .

NIELSEN, F. (2018): "What is an information projection," *Notices of the AMS*, 65, 321–324.

RUBINSTEIN, M. (1974): "An aggregation theorem for securities markets," *Journal of Financial Economics*, 1, 225–244.

SARGENT, T. (1976): "Econometric exogeneity and alternative estimators of portfolio balance schedules for hyperinflations: A note," *Journal of Monetary Economics*, 2, 511–521.

SARGENT, T. J. (2015): "Robert E. Lucas Jr.'s Collected Papers on Monetary Theory," *Journal of Economic Literature*, 53, 43–64.

——— (2024): "HAOK and HANK Models," in *Heterogeneity in Macroeconomics: Implications for Monetary Policy*, ed. by S. Bauducco, A. Fernández, and G. L. Violante, Santiago, Chile: Central Bank of Chile, vol. 30 of *Series on Central Banking, Analysis, and Economic Policies*, 13–38.

SARGENT, T. J., Y. J. SELVAKUMAR, AND Z. YANG (2025): "Aggregate Shocks and Cross-Section Dynamics: Quantifying Redistribution and Insurance in US Household Data," .

SARGENT, T. J. AND C. A. SIMS (1977): "Business cycle modeling without pretending to have too much a priori economic theory," in *New methods in business cycle research*, Minneapolis, Minnesota: Federal Reserve Bank of Minneapolis, 145–168.

SARGENT, T. J. AND J. STACHURSKI (2025): "Gorman Aggregation," https://python-advanced.quantecon.org/gorman_heterogeneous_households.html, advanced Quantitative Economics with Python lecture, QuantEcon.

STOCK, J. H. AND M. W. WATSON (2002): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

——— (2016): "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," *Handbook of Macroeconomics*, 2, 415–525.

STRANG, G. (2020): *Linear Algebra and Learning from Data*, Wellesley, MA: Cambridge University Press.

TU, J. H., C. W. ROWLEY, AND D. M. LUCHTENBURG (2014): "On Dynamic Mode Decomposition: Theory and Application," *Journal of Computational Dynamics*, 1, 391–421.

WHITE, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–25.

WILLIAMS, M. O., I. G. KEVREKIDIS, AND C. W. ROWLEY (2015): "A Data–Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition," *Journal of Nonlinear Science*, 25, 1307–1346.