Dynamic Mode Decompositions and Vector Autoregressions *

Thomas J. Sargent
New York University

Yatheesan J. Selvakumar
New York University

October 2, 2024

Abstract

Sufficient conditions on state-space matrices $\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R}$ allow inferring them from a reduced-rank first-order vector autoregression (VAR) that can be computed with a Dynamic Mode Decomposition (DMD), thereby connecting DMD modes to hidden Markov states in the state-space system. When these sufficient conditions hold, our technique provides a fast way to infer parameters of the linear state space system. An application infers two macroeconomic aggregate factors from dynamics of CEX cross sections.

Keywords: Vector Autoregression, LGQ State-Space Model, Innovations Representation, Singular Value Decomposition, Dynamic Mode Decomposition,

*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 Cast of Characters

Brunton and Kutz (2022) describe applications of Dynamic Mode Decompositions (DMDs) to a variety of problems. Applied mathematicians have recently used them as a "machine learning" technique to construct deterministic dynamical systems, including fluid dynamics. This paper applies DMDs to a dynamical system buffeted by random shocks. We analyze connections among Dynamic Mode Decompositions (DMDs), vector autoregressions (VARs), and a special case of what Stock and Watson (2016) call dynamic factor models. At critical points we deploy singular value decompositions and associated eigen decompositions, but we apply them in different ways and to different objects than do the papers surveyed by Stock and Watson.[1]

Our paper focuses on five objects:

1. A hidden Markov Model for a covariance stationary stochastic process $\{\mathbf{y}_t\}_{t=-\infty}^{\infty}$ that takes the form of an LQG state-space model, where $\mathbf{y}_t$ is an $M \times 1$ vector.
2. An innovations representation of the hidden Markov 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 $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T, \mathbf{y}_{T+1}]$

We describe 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]

[1] See Appendix A for details.

The infinite-order VAR in item [3] is

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

$$\mathbb{E}[\mathbf{a}_t \mathbf{y}_{t-j}^{\top}] = \mathbf{0} \quad \text{for all } j \geq 1$$

where $\mathbb{E}[\mathbf{a}_t \mathbf{a}_t^T] = \mathbf{\Omega}$ for all t. We reverse engineer circumstances in which

- $\mathbf{B}_j^{\infty} = 0 \quad \forall j > 1$
- $\mathbf{B}_1^{\infty}$ is an $M \times M$ matrix of rank $N << M$,

so that the infinite-order VAR becomes a reduced-rank first-order VAR. We connect that reduced-rank first-order VAR to a time-invariant innovations representation

$$\widehat{\mathbf{x}}_{t+1} = \mathbf{A} \ \widehat{\mathbf{x}}_t + \mathbf{K} \, \mathbf{a}_t$$

$$\mathbf{y}_t = \mathbf{G} \ \widehat{\mathbf{x}}_t + \mathbf{a}_t, \ \ \mathbf{a}_t \perp \mathbf{y}^{t-1},$$

where $\mathbf{A}$ is $N \times N$, $\mathbf{G}$ is $M \times N$, and $\mathbf{K}$ is $M \times N$, $\mathbf{y}^{t-1} = \{\mathbf{y}_{t-s}\}_{s=1}^{\infty}$, $\perp$ indicates orthogonality, and where the innovations representation is associated with an LQG state space representation for the $\{\mathbf{y}_t\}_{t=-\infty}^{\infty}$ process:

$$\mathbf{x}_{t+1} = \mathbf{A} \, \mathbf{x}_t + \mathbf{C} \, \mathbf{w}_{t+1}$$

$$\mathbf{y}_t = \mathbf{G} \, \mathbf{x}_t + \mathbf{v}_t,$$

where random shocks $\mathbf{w}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N \times N})$, measurement errors $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{M \times M})$ and $\mathbf{w}_s \perp \mathbf{v}_\tau$ for all s, τ; here $\mathbf{C}$ is $N \times N$.

Connections among the five objects are intermediated through a Dynamic Mode Decomposition (DMD) that we use to infer a reduced-rank first-order VAR from a data set $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T, \mathbf{y}_{T+1}]$.[2] We use these connections to infer the matrices $\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R}$ of the

[2]Through its connection to the Koopman operator, the DMD algorithm has also been used to approximate non-linear dynamics (see Brunton et al. (2016), Mezic (2020) and Williams et al. (2015)).

hidden Markov model and also the Kalman gain matrix $\mathbf{K}$ in the associated innovations representation from a reduced-rank first-order vector autoregression (VAR).[3] When applicable, this machinery provides a good way to estimate parameters of dynamic stochastic equilibrium models with large numbers of endogenous variables (Iao and Selvakumar (2024) offer examples).

Section 2 applies a Dynamic Mode Decomposition (DMD) algorithm of Tu et al. (2014) to estimate a reduced-rank first-order vector autoregression in a setting in which least-squares regression coefficients are underdetermined.[4] Section 2 also describes DMD "dynamic modes" and how to compute them. Section 3 relates dynamic modes to objects in a "pseudo innovations representation" that approximates an authentic innovations representation that is affiliated with a linear state-space model.[5] We state restrictions on the linear state space model that align the infinite-order VAR implied by an authentic innovations representation with the reduced-rank first-order VAR associated with the DMD algorithm. Under these restrictions, DMD modes estimate hidden Markov states (a.k.a. "dynamic factors") in that state-space model. The restrictions make each hidden factor in the state-space model follow a univariate first-order autoregression with shocks that are possibly correlated contemporaneously across factors.[6] As the number of variables M in the observation vector grows, a pseudo innovations representation aligns better and better with an authentic innovations representation. This outcome implies that a large enough and sufficiently independent

[3] Our approach is in the spirit of distinct statistical models proposed by 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 them, we estimate "non-structural Kepler stage" descriptive models that can compress data and reveal patterns. As Koopmans (1947) recommends, these empirical regularities are subsequently to be interpreted by "structural Newton stage" models 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). Iao and Selvakumar (2024) apply our techniques to estimate an auxiliary model that they then use as part of a strategy for estimating structural parameters of what purports to be a "Newton stage" modern macro model with many heterogeneous agents.

[4] Brunton and Kutz (2022, sec. 7.2) describes the algorithm.

[5] Ljungqvist and Sargent (2018, secs. 2.7-2.10) describe innovations representations and link them to vector autoregressions.

[6] Appendix A compares identification restrictions imposed in various dynamic factor models.

collection of contemporaneous noisy measurements of linear combinations of the hidden factors contains as much information as would be provided by an infinite history of those measurements.[7] When these conditions prevail, matrices defining a state-space model can be recovered from objects computed by DMD – a finding that justifies an inexpensive algorithm for estimating parameters of a state-space system. Section 4 illustrates, checks, and stress-tests our section 3 theoretical findings by conducting two "lab experiments" in the context of an example theoretical model. Section 5 uses a reduced-rank first-order VAR to describe dynamics of US CEX cross sections for quarterly US data from 1990 to 2021. We detect and venture to label two dominant DMD modes, one that we interpret as a macroeconomic "reference cycle" in the tradition of Burns and Mitchell (1946), as reinterpreted by Koopmans (1947), and another that we interpret as an "inequality mode" in the spirit of recent empirical macroeconomic research in the spirit of Guvenan et al. (2014, 2017). Section 6 offers concluding remarks that point to how Iao and Selvakumar (2024) have used a reduced-rank VAR as an auxiliary model that helps them infer parameters of a heterogeneous agent dynamic general equilibrium model.

2 Reduced-Rank First-Order VAR

2.1 Data and Statistical Model

Let $\mathbf{y}_t$ be an an $M \times 1$ vector of de-meaned random variables at $t = 1, \ldots, T+1$, and assume that $M > T$, so that there are more variables than time periods.[8] Horizontally stack $\mathbf{y}_t$ vectors across times to create two $M \times T$ "tall and skinny" data matrices $\mathbf{Y}$

[7] That averaging observations over large cross-sections accelerates learning about hidden dynamic factors is reminiscent of Chamberlain and Rothschild (1982). More recent examples include Forni et al. (2000), Forni and Lippi (2001), and a variety of the papers surveyed in Stock and Watson (2016).

[8] 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.

and $\mathbf{Y}'$:

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T] \tag{1}$$

$$\mathbf{Y}' = [\mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_{T+1}]. \tag{2}$$

We want to estimate a first-order vector autoregression

$$\mathbf{y}_t = \mathbf{B}\,\mathbf{y}_{t-1} + \mathbf{a}_t\,, \quad \mathbf{a}_t \perp \mathbf{y}_{t-1} \tag{3}$$

$$\mathbb{E}[\mathbf{a}_t\,\mathbf{a}_t^\top] = \mathbf{\Omega}, \tag{4}$$

where $\perp$ indicates orthogonality. We have $M(T+1)$ data points for estimating M^2 parameters in $\mathbf{B}$. Since $M^2 > M(T+1)$, the least squares estimator $\widehat{\mathbf{B}}$ of $\mathbf{B}$ is underdetermined, so we choose

$$\widehat{\mathbf{B}} = \underset{\text{rank}(B)=N}{\arg\min} \,||\,\mathbf{Y}' - B\,\mathbf{Y}\,||_F \tag{5}$$

where $||\,\mathbf{D}\,||_F$ denotes the Frobenius norm $\sum_{i,j}\mathbf{D}_{i,j}^2$ of matrix D. To compute $\widehat{\mathbf{B}}$, we first represent $\mathbf{Y}$ with a reduced Singular Value Decomposition (SVD)[9]

$$\mathbf{Y} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^\top,$$

where $\widetilde{\mathbf{U}}$ is $M \times T$, $\widetilde{\mathbf{\Sigma}}$ is $T \times T$ and $\widetilde{\mathbf{V}}^\top$ is $T \times T$. We compress $\mathbf{Y}$ by using the N largest singular values to form

$$\mathbf{Y} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top, \tag{6}$$

where $\mathbf{U} = \widetilde{\mathbf{U}}[:, :N]$, $\mathbf{\Sigma} = \widetilde{\mathbf{\Sigma}}[:N, :N]$ has N singular values as its only non-zero entries, and $\mathbf{V}^\top = \widetilde{\mathbf{V}}^\top[:N, :]$. Here $\mathbf{U}$ is $M \times T$, $\mathbf{V}$ is $T \times N$, $\mathbf{\Sigma}$ is $N \times N$, and $\mathbf{V}^\top$ is $N \times T$. From now on we let $\mathbf{Y}$ presented in approximation (6) denote the compressed version

[9] Anderson (1951) and Anderson (1999) estimate reduced-rank regressions by first computing an unrestricted least-squares regression. This approach is infeasible in our $M > T$ setting.

of the original data set. We use the reduced SVD approximation (6) of the original $\mathbf{Y}$ matrix as an input to computing

$$\widehat{\mathbf{B}} = \mathbf{Y}' \, \mathbf{Y}^{+},$$

where $\mathbf{Y}^{+} = \mathbf{V} \, \mathbf{\Sigma}^{-1} \, \mathbf{U}^{\top}$ is a generalized inverse of $\mathbf{Y}$ that verifies $\mathbf{Y}^{+} \, \mathbf{Y} = \mathbf{I}_{T \times T}$ and $\mathbf{Y}'$ is defined in (2).

We use these inputs to represent the reduced-rank first-order VAR (3) in terms of N "dynamic modes" by implementing the following steps:[10]

1. Use $\mathbf{U}$ and $\widehat{\mathbf{B}}$ to construct $\widetilde{\mathbf{B}}_{N \times N} = \mathbf{U}^{\top} \, \widehat{\mathbf{B}} \, \mathbf{U}$.

2. Construct an eigendecomposition

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

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

3. Define

$$\mathbf{\Phi}_{M \times N} \equiv \mathbf{Y}' \, \mathbf{V} \, \mathbf{\Sigma}^{-1} \, \mathbf{W} \, .$$

4. Take advantage of a finding of Tu et al. (2014) that columns of $\mathbf{\Phi}$ are eigenvectors of $\widehat{\mathbf{B}}$ that share eigenvalues with $\widetilde{\mathbf{B}}$, so that

$$\widehat{\mathbf{B}} = \mathbf{\Phi} \mathbf{\Lambda} \mathbf{\Phi}^{+}, \tag{7}$$

   where $\mathbf{\Phi}$ is $M \times N$, $\mathbf{\Lambda}$ is $N \times N$, and $\mathbf{\Phi}^{+}$ is the $N \times M$ (left) generalized inverse that verifies $\mathbf{\Phi}^{+} \mathbf{\Phi} = I_{N \times N}$.

[10] Here we are following and extending steps described by Brunton and Kutz (2022, sec. 7.2).

5. Define an $N \times 1$ vector of "dynamic modes" by[11]

$$\widetilde{\mathbf{x}}_t \equiv \mathbf{\Phi}^+ \mathbf{y}_t \,.$$

6. Use representation (7) of $\widehat{\mathbf{B}}$ to express the reduced rank first-order VAR as

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^+ \mathbf{y}_{t-1} + \widehat{\mathbf{a}}_t, \quad \widehat{\mathbf{a}}_t \perp \mathbf{y}_{t-1} \tag{8}$$

7. Compute

$$\widehat{\mathbf{\Omega}} = \frac{1}{T-1} \sum_{t=1}^{T} \widehat{\mathbf{a}}_t \widehat{\mathbf{a}}_t^\top \tag{9}$$

8. Multiply both sides of equation (8) by the $N \times M$ matrix $\mathbf{\Phi}^+$ to obtain $\widetilde{\mathbf{x}}_t = \mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1} + \mathbf{\Phi}^+\widehat{\mathbf{a}}_t$. Use it together with $\widetilde{\mathbf{x}}_{t-1} = \mathbf{\Phi}^+ \mathbf{y}_{t-1}$ in equation (8) to form the system

$$\widetilde{\mathbf{x}}_t = \mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1} + \mathbf{\Phi}^+\widehat{\mathbf{a}}_t \tag{10}$$

$$\mathbf{y}_t = \mathbf{\Phi}\mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1} + \widehat{\mathbf{a}}_t, \tag{11}$$

where $\mathbf{\Phi}\mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1} = \widehat{\mathbf{B}}\,\mathbf{y}_{t-1}$ and $\widehat{\mathbf{a}}_t \perp \mathbf{y}_{t-1}$ but $\widehat{\mathbf{a}}_t$ is not necessarily orthogonal to history $\mathbf{y}^{t-1}$, i.e., $\widehat{\mathbf{a}}_t \not\perp \mathbf{y}^{t-1}$.

9. Construct a moving average representation for modes $\widetilde{\mathbf{x}}_t$:

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

10. Construct j step ahead conditional covariances of modes:

[11] In section 3, we'll link these modes to an $N \times 1$ hidden state vector $\mathbf{x}_t$ that is defined implicitly by $\widetilde{\mathbf{x}}_t \equiv \mathbb{E}[\mathbf{x}_t \,|\, \mathbf{y}_t]$. We'll also define a distinct projection $\widehat{\mathbf{x}}_t = \mathbb{E}[\mathbf{x}_t \,|\, \mathbf{y}_{t-1}]$.

$$\mathbb{E}(\widetilde{\mathbf{x}}_{t+j} - \mathbb{E}\,\widetilde{\mathbf{x}}_{t+j}|\widetilde{\mathbf{x}}_t)(\mathbf{x}_{t+j} - \mathbb{E}\,\widetilde{\mathbf{x}}_{t+j}|\widetilde{\mathbf{x}}_t)^\top = \sum_{s=0}^{j-1} \mathbf{\Lambda}^s \mathbf{\Phi}^+ \widehat{\mathbf{\Omega}} (\mathbf{\Phi}^+)^\top \mathbf{\Lambda}^s \tag{12}$$

In sections 3 and 4, we describe conditions under which outcomes of the DMD algorithm are associated with objects that define a linear, Gaussian hidden Markov model. Those sections also indicate other purposes a DMD analysis can serve.

3 Two Innovations Representations

In section 2, we described how to construct a reduced-rank first-order vector autoregression and then use it to cast representation (10)-(11) in terms of dynamic modes $\widetilde{\mathbf{x}}_t$. In this section, we use equations (10)-(11) to construct what we call a **pseudo innovations representation** cast in terms of an $N \times 1$ vector $\widehat{\mathbf{x}}_t$. To accomplish this, we begin by recalling that in section 2 we implicitly defined an $N \times 1$ vector $\mathbf{x}_t$ by a projection

$$\widetilde{\mathbf{x}}_t = \mathbb{E}\,\mathbf{x}_t \,|\, \mathbf{y}_t\,.$$

Define a distinct projection

$$\widehat{\mathbf{x}}_t = \mathbb{E}\,\mathbf{x}_t \,|\, \mathbf{y}_{t-1}\,.$$

System (10)-(11) implies that $\widehat{\mathbf{x}}_t$ is related to $\widetilde{\mathbf{x}}_{t-1}$ by

$$\widehat{\mathbf{x}}_t = \mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1} = \mathbf{\Lambda}\mathbf{\Phi}^+\,\mathbf{y}_{t-1}\,. \tag{13}$$

With equation (13) in mind, multiply both sides of (10) by $\mathbf{\Lambda}$ and substitute $\widehat{\mathbf{x}}_t$ for $\mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1}$ in equation (11) to obtain the **pseudo innovations representation**:

$$\widehat{\mathbf{x}}_{t+1} = \mathbf{\Lambda}\widehat{\mathbf{x}}_t + \mathbf{\Lambda}\mathbf{\Phi}^+\widehat{\mathbf{a}}_t \tag{14}$$

$$\mathbf{y}_t = \mathbf{\Phi}\widehat{\mathbf{x}}_t + \widehat{\mathbf{a}}_t, \;\; \widehat{\mathbf{a}}_t \perp \mathbf{y}_{t-1}\,. \tag{15}$$

It is enlightening to compare representation (14)-(15) with an authentic time-invariant innovations representation

$$\widehat{\mathbf{x}}_{t+1} = \mathbf{A}\;\widehat{\mathbf{x}}_t + \mathbf{K}\,\mathbf{a}_t \tag{16}$$

$$\mathbf{y}_t = \mathbf{G}\;\widehat{\mathbf{x}}_t + \mathbf{a}_t, \;\; \mathbf{a}_t \perp \mathbf{y}^{t-1} \tag{17}$$

that is associated with the linear state-space model

$$\mathbf{x}_{t+1} = \mathbf{A}\,\mathbf{x}_t + \mathbf{C}\,\mathbf{w}_{t+1} \tag{18}$$

$$\mathbf{y}_t = \mathbf{G}\,\mathbf{x}_t + \mathbf{v}_t, \tag{19}$$

where the $N \times M$ matrix $\mathbf{K}$ in equation (16) is the steady-state Kalman gain associated with (18)-(19), shocks $\mathbf{w}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N\times N})$, measurement errors $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{M\times M})$ and $\mathbf{w}_s \perp \mathbf{v}_\tau$ for all s, τ; here $\mathbf{A}$ is $N \times N$, $\mathbf{C}$ is $N \times N$ and $\mathbf{G}$ is $M \times N$. Now $\widehat{\mathbf{x}}_t = \mathbb{E}[\mathbf{x}_t \,|\, \mathbf{y}^{t-1}]$, and $\mathbf{a}_t = \mathbf{y}_t - \mathbb{E}[\mathbf{y}_t \,|\, \mathbf{y}^{t-1}]$, $\mathbf{a}_t \perp \mathbf{a}_s \,\forall t \neq s$ for $\mathbf{y}^t = \{\mathbf{y}_s\}_{s<t}$ and $\mathcal{H}(\mathbf{a}^t) = \mathcal{H}(\mathbf{y}^t)$ connect Hilbert spaces spanned by histories $\mathbf{a}^t$ and $\mathbf{y}^t$, respectively.[12] Furthermore,

$$\mathbf{\Omega} \equiv \mathbb{E}[\mathbf{a}_t\,\mathbf{a}_t^\top] = \mathbf{G}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top + \mathbf{R},$$

[12]The notation H($\mathbf{y}^t$) indicates a space spanned by all square-summable linear combinations of history $\mathbf{y}^t$, and so on.

where $\mathbf{\Sigma}_\infty$ is the positive semi-definite solution of the algebraic matrix Riccati equation

$$\begin{aligned}\mathbf{\Sigma}_\infty &= \mathbb{E}[\mathbf{x}_t - \widehat{\mathbf{x}}_t][\mathbf{x}_t - \widehat{\mathbf{x}}_t]^\top \\ &= \mathbf{C}\,\mathbf{C}^\top + \mathbf{K}\,\mathbf{R}\,\mathbf{K}^\top + (\mathbf{A} - \mathbf{K}\,\mathbf{G})\mathbf{\Sigma}_\infty(\mathbf{A} - \mathbf{K}\,\mathbf{G})^\top \end{aligned} \tag{20}$$

and the steady-state Kalman gain matrix $\mathbf{K}$ satisfies

$$\mathbf{K} = \mathbf{A}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top(\mathbf{G}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top + \mathbf{R})^{-1}. \tag{21}$$

Comparing system (14)-(15) with system (16)-(17) tempts us to match objects according to

$$\mathbf{A} = \mathbf{\Lambda}, \quad \mathbf{K} = \mathbf{\Lambda}\mathbf{\Phi}^+, \quad \mathbf{G} = \mathbf{\Phi} \tag{22}$$

However, the distinct least-squares orthogonality conditions $\widehat{\mathbf{a}}_t \perp \mathbf{y}_{t-1}$ and $\mathbf{a}_t \perp \mathbf{y}^{t-1}$ are tell-tale signs that the pseudo-innnovations representation (14)-(15) is associated with a **first-order** VAR, while an authentic innovations representation (16)-(17) is associated with an **infinite-order** VAR. This means that it is not appropriate to expect the mapping in (22) to prevail in general.

However, in subsection 3.1, we describe restrictions on the linear state space model (18)-(19) that make a first-order VAR satisfy least-squares orthogonality conditions that promote it to an infinite-order VAR, thereby activating connections (22).[13] Under those restrictions we can infer parameters of a linear state space model (18)-(19) from our reduced-dimension first-order VAR objects $\mathbf{\Lambda}, \mathbf{\Phi}$.

3.1 Restrictions on State-Space Model

We impose the following restrictions on state-space model (18)-(19):

[13] This requires that the innovations $\widehat{\mathbf{a}}_t$ be orthogonal to the linear space spanned by $\mathbf{y}^{t-1}$ and not just orthogonal to $\mathbf{y}_{t-1}$.

1. $M >> N$

2. $\mathbf{A}$ is a diagonal

3. $\mathbf{G}$ has full column rank

4. $||\mathbf{G}^\top \mathbf{G}|| = \mathcal{O}(M)$ and $||\mathbf{R}|| = o(M)$ where $||\cdot||$ denotes the Frobenius norm

Item 1 asserts that cross autocorrelations among components of $\{\mathbf{y}_t\}$ are intermediated by common dependencies on N hidden state variables. Item 2 asserts that each component of $\mathbf{x}_t$ follows an AR(1) process with shocks that can be correlated contemporaneously across components. Item 3 requires columns of $\mathbf{G}$ to be linearly independent.[14,15] Item 4 requires that, as the number of observables becomes large, $||\,\mathbf{G}^\top\,\mathbf{G}\,||$ does not grow "too fast". The item 4 restriction on $\mathbf{R}$ allows for heteroskedasticity in the measurement errors, but requires that variances cannot grow faster than rate M. Stock and Watson (1988) describe a related but stronger assumption that requires that the maximum eigenvalue of $\mathbf{R}$ be bounded, which limits cross-correlations between measurement errors.[16,17]

To help connect pseudo and authentic innovation representations, we recycle notation and temporarily define $\widetilde{\mathbf{x}}_t \equiv \mathbb{E}[\mathbf{x}_t \,|\, \mathbf{y}^t]$, so that now

$$\widetilde{\mathbf{x}}_t = \widehat{\mathbf{x}}_t + \mathbf{L}\widehat{\mathbf{a}}_t \tag{23}$$

where

$$\mathbf{L} = \mathbf{\Sigma}_\infty \,\mathbf{G}^\top (\mathbf{G}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top + \mathbf{R})^{-1}, \tag{24}$$

[14] Item 3 is an identifying assumption on $\mathbf{G}$ that precludes rewriting the state-space model with fewer factors. To see this, suppose $\mathbf{G}$ was rank $N-1$. Then it would be possible to rewrite the linear state-space model with $N-1$ factors, implying that only $N-1$ factors are identified.

[15] It is useful to compare our restrictions with those coming from a principle components analysis of Bai and Ng (2008) who describe a set of possible identifying restrictions on $\mathbf{G}$ and $\mathbf{x}_t$ that involve either (a) orthogonality of the hidden factors, or (b) orthogonality of the loadings $\mathbf{G}$, and/or (c) some zero restrictions on elements of $\mathbf{G}$. By contrast, item 3 does not restrict any particular elements of $\mathbf{G}$.

[16] Item 4 is also a common assumption in the principal components literature; see for example Bai and Ng (2013) and Chamberlain and Rothschild (1983).

[17] Iao and Selvakumar (2024) also place these assumptions on a linear state-space model to estimate structural parameters of heterogeneous-agent models with high-dimensional micro data.

and consequently

$$\widetilde{\mathbf{x}}_{t+1} = \mathbf{A}\,\widetilde{\mathbf{x}}_t + \mathbf{\Phi}^{+}\widehat{\mathbf{a}}_{t+1} \tag{25}$$

$$\mathbf{y}_{t+1} = \mathbf{\Phi}\mathbf{\Lambda}\widetilde{\mathbf{x}}_t + \widehat{\mathbf{a}}_{t+1} \tag{26}$$

3.2 Connections

Innovations representation (16)-(17) is associated with an infinite-order VAR for the $\{\mathbf{y}_t\}$ process:

$$\mathbf{y}_t = \sum_{j=1}^{\infty} \mathbf{B}_j^{\infty}\,\mathbf{y}_{t-j} + \mathbf{a}_t \tag{27}$$

$$\mathbb{E}[\mathbf{a}_t\,\mathbf{y}_{t-j}^{\top}] = \mathbf{0} \quad \text{for all } j \geq 1$$

$$\mathbb{E}[\mathbf{a}_t\,\mathbf{a}_t^{T}] = \mathbf{\Omega} = \mathbf{G}\,\mathbf{\Sigma}_{\infty}\,\mathbf{G}^{\top} + \mathbf{R} \tag{28}$$

$$\mathbf{B}_j^{\infty} = \mathbf{G}(\mathbf{A} - \mathbf{K}\,\mathbf{G})^{j-1}\,\mathbf{K} \quad \forall j \geq 1 \tag{29}$$

where $\text{rank}(\mathbf{B}_j^{\infty}) = N \quad \forall j \geq 1$. Under the four subsection 3.1 restrictions on state-space system (18)-(19), $(\mathbf{A} - \mathbf{K}\,\mathbf{G}) \approx \mathbf{0}$, so that (29) implies[18]

$$\mathbf{B}_j^{\infty} \approx \begin{cases} \mathbf{G}\,\mathbf{K} & j = 1 \\ \mathbf{0} & j > 1 \end{cases} \tag{30}$$

Consequently, the infinite order VAR (27) becomes a first order VAR

$$\mathbf{y}_t = \mathbf{B}_1^{\infty}\,\mathbf{y}_{t-1} + \mathbf{a}_t, \tag{31}$$

which is equivalent to the reduced-rank first-order VAR (3) with $\mathbf{B} \approx \mathbf{B}_1^{\infty}$. This means that the pseudo-innovations representation approximates an innovations representa-

[18] See Iao and Selvakumar (2024) for a proof.

tion well and rationalizes equations (22). These equalities allow us to infer $\mathbf{R}$ and $\mathbf{C}\,\mathbf{C}^\top$. From the formulas (21) and (20) for $\mathbf{K}$ and $\mathbf{\Sigma}_\infty$, respectively, it follows that

$$\mathbf{A}^\top - \mathbf{G}^\top\,\mathbf{K}^\top = \left[\mathbf{I} - \mathbf{G}^\top\,\mathbf{\Omega}^{-1}\,\mathbf{G}\,\mathbf{\Sigma}_\infty\right]\mathbf{A}^\top\,.$$

When $\mathbf{A} - \mathbf{K}\,\mathbf{G} = \mathbf{0}$, this formula implies that $\mathbf{I} = \mathbf{G}^\top\,\mathbf{\Omega}^{-1}\,\mathbf{G}\,\mathbf{\Sigma}_\infty$, which implies that $\mathbf{\Sigma}_\infty$ satisfies

$$\mathbf{\Sigma}_\infty = (\mathbf{G}^\top\,\mathbf{\Omega}^{-1}\,\mathbf{G})^{-1}.$$

Then from equation from (28), we infer

$$\mathbf{R} = \mathbf{\Omega} - \mathbf{G}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top\,. \tag{32}$$

When $\mathbf{A} - \mathbf{K}\,\mathbf{G} = \mathbf{0}$, equation (20) implies $\mathbf{\Sigma}_\infty = \mathbf{C}\,\mathbf{C}^\top + \mathbf{K}\,\mathbf{R}\,\mathbf{K}^\top$, which in turn implies

$$\mathbf{C}\,\mathbf{C}^\top = \mathbf{\Sigma}_\infty - \mathbf{K}\,\mathbf{R}\,\mathbf{K}^\top\,. \tag{33}$$

Next, solve the discrete Lyapunov equation

$$\mathbf{V}_\mathbf{x} = \mathbf{A}\,\mathbf{V}_\mathbf{x}\,\mathbf{A}^\top + \mathbf{C}\,\mathbf{C}^\top \tag{34}$$

affiliated with state evolution equation (18) for $\{\mathbf{x}_t\}$, the stationary covariance matrix of hidden state vector $\mathbf{x}_t$. Then we can compute the stationary covariance matrix $\mathbf{V}_\mathbf{y}$ of the measurement process $\{\mathbf{y}_t\}$ from

$$\mathbf{V}_\mathbf{y} = \mathbf{G}\,\mathbf{V}_\mathbf{x}\,\mathbf{G}^\top + \mathbf{R}\,. \tag{35}$$

3.3 Algorithm

Algorithm 1 provides pseudo-code for estimating $\mathbf{A}, \mathbf{C}\,\mathbf{C}^\top, \mathbf{G}, \mathbf{R}$ state-space model (18)-(19). We apply the algorithm to a suite of laboratory examples in the section 4.[19]

Algorithm 1 Pseudo-code for inferring $\mathbf{A}, \mathbf{G}, \mathbf{R}, \mathbf{C}\,\mathbf{C}^\top$

1. Set number of modes N
2. Compute $\mathbf{\Phi}$, $\mathbf{\Lambda}$ and $\widehat{\mathbf{B}}$ via DMD.
3. Approximate $\mathbf{A}$ with $\mathbf{\Lambda}$.
4. Approximate $\mathbf{G}$ with $\mathbf{\Phi}$.
5. Approximate $\mathbf{L}$ with $\mathbf{\Phi}^+$ and $\mathbf{K}$ with $\widehat{\mathbf{K}} = \mathbf{\Lambda}\mathbf{\Phi}^+$
6. Approximate $\widetilde{\mathbf{x}}_t$ with $\mathbf{\Phi}^+\,\mathbf{y}_t$.
7. Approximate $\widehat{\mathbf{x}}_t$ with $\mathbf{\Lambda}\widetilde{\mathbf{x}}_{t-1}$.
8. Approximate $\mathbb{E}[\mathbf{y}_{t+j}\,|\,\mathbf{y}_t] = \mathbf{G}\,\mathbf{A}^j\,\widetilde{\mathbf{x}}_t$ with $\mathbf{\Phi}\mathbf{\Lambda}^j\widetilde{\mathbf{x}}_t$.
9. Approximate $\mathbf{\Omega}$ with
$$\widehat{\mathbf{\Omega}} = \frac{1}{T-1}\sum_{t=1}^{T}\widehat{\mathbf{a}}_t\widehat{\mathbf{a}}_t^\top$$
where $\widehat{\mathbf{a}}_t = \mathbf{y}_t - \widehat{\mathbf{B}}\,\mathbf{y}_{t-1}$.
10. Approximate $\mathbf{\Sigma}_\infty$ with
$$\widehat{\mathbf{\Sigma}}_\infty = (\mathbf{\Phi}^\top\widehat{\mathbf{\Omega}}^{-1}\mathbf{\Phi})^{-1}$$
11. Approximate $\mathbf{R}$ with
$$\widehat{\mathbf{R}} = \widehat{\mathbf{\Omega}} - \mathbf{\Phi}\widehat{\mathbf{\Sigma}}_\infty\mathbf{\Phi}^\top$$
12. Approximate $\mathbf{C}\,\mathbf{C}^\top$ with
$$\widehat{\mathbf{C}\,\mathbf{C}^\top} = \widehat{\mathbf{\Sigma}}_\infty - \widehat{\mathbf{K}}\widehat{\mathbf{R}}\widehat{\mathbf{K}}^\top$$

[19] Appendix B discusses a computational detail in inverting rank-deficient matrices.

4 Laboratory

In this section, we apply our algorithm to outcomes generated by a suite of state-space models with two factors $N = 2$. Models differ in the length M of the measurement vector $\mathbf{y}_t$. We set the state-space matrices according to

$$\mathbf{A} = \begin{bmatrix} 0.9 & 0 \\ 0 & 0.7 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0.5 & 0.4 \\ 0 & 0.5 \end{bmatrix} \tag{36}$$

$$\widetilde{\mathbf{G}} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}, \quad \mathbf{R} = 0.25\mathbf{I}_M$$

The first $\frac{M}{2}$ rows of $\mathbf{G}$ are $[1, 0]$ and the second $\frac{M}{2}$ rows are $[0, 1]$, so $\mathbf{G}$ has full column rank.

We make two comparisons. The first compares sample outcomes delivered by Algorithm 1 with the population objects implied by the matrices in equation (36) (see Section 4.1). The second studies finite-sample biases of the DMD implementation by comparing averages of statistics from repeated random samples with the same population objects (see Section 4.2).

4.1 Population Objects

From $\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{R}$ we compute population objects $\mathbf{B}, \mathbf{K}, \mathbf{\Sigma}_\infty, \mathbf{\Omega}, \mathbf{B}_1^\infty$ by first enlisting the quantecon **LinearStateSpace** class to compute population moments of the stationary distribution of the process $\{\mathbf{x}_t, \mathbf{y}_t\}$ and the associated population cross-covariagrams

$\mathbb{E}[\mathbf{x}_t \mathbf{x}_t^\top] \equiv \mathbf{\Sigma_x}$, $\mathbb{E}[\mathbf{y}_t \mathbf{y}_t^\top] \equiv \mathbf{\Sigma_y}$, and $\mathbb{E}[\mathbf{x}_t \mathbf{y}_t^\top] \equiv \mathbf{\Sigma_{xy}}$. These allow us to compute population first-order autoregressive coefficients $\mathbf{B}$ via

$$\mathbf{B} = \mathbf{G}\,\mathbf{A}\,\mathbf{\Sigma_x}\,\mathbf{G}^\top\,\mathbf{\Sigma_y}^{-1}$$

After that, we use the quantecon class **Kalman** to compute the innovations representation and the associated $\mathbf{K}, \mathbf{\Sigma}_\infty$. This allows us to compute the population $\mathbf{B}_1^\infty$ and one-step-ahead conditional covariance matrix $\mathbf{\Omega}$ via

$$\begin{aligned}\mathbf{B}_1^\infty &= \mathbf{G}\,\mathbf{K} \\ \mathbf{\Omega} &= \mathbf{G}\,\mathbf{\Sigma}_\infty\,\mathbf{G}^\top + \mathbf{R}\,.\end{aligned}$$

Next, we use equation (32) to calculate a population analogue of $\widehat{\mathbf{R}}$ by setting $\mathbf{G} = \mathbf{\Phi}$. We then use $\widehat{\mathbf{R}}$ as an input to step **12** from Pseudo-code 1 to compute a population analogue of $\widehat{\mathbf{C}\,\mathbf{C}}^\top$.

We perform these calculations for $M = 2, M = 300, M = 1000$ to obtain outcomes that we report in Table 1. Frobenius norms reported in the first three rows describe how well a pseudo innovations representation approximates an authentic one. As anticipated, as M gets larger, $||\,\mathbf{A} - \mathbf{K}\,\mathbf{G}\,||_F$ approaches zero, as do approximation error measures $M^{-1}||\,\mathbf{B} - \mathbf{B}_1^\infty\,||_F$ and $M^{-1}||\,\mathbf{K} - \mathbf{A}\,\mathbf{G}^+\,||_F$[20]. In addition, $\widehat{\mathbf{C}\,\mathbf{C}^\top}$ approaches $\mathbf{C}\,\mathbf{C}^\top$ and $\widehat{\mathbf{R}}$ approaches $\mathbf{R}$.

Object	$M = 2$	$M = 300$	$M = 1000$
$\|\|\,\mathbf{A} - \mathbf{K}\,\mathbf{G}\,\|\|_F$	0.5	0.02	0.003
$M^{-1}\|\|\,\mathbf{B} - \mathbf{B}_1^\infty\,\|\|_F$	0.2	$3e^{-5}$	$3e^{-6}$
$M^{-1}\|\|\,\mathbf{K} - \mathbf{A}\,\mathbf{G}^+\,\|\|_F$	0.25	$3e^{-6}$	$2e^{-7}$
$M^{-1}\|\|\widehat{\mathbf{R}} - \mathbf{R}\,\|\|_F$	0.2	0.001	0.0004
$\|\|\widehat{\mathbf{C}\,\mathbf{C}^\top} - \mathbf{C}\,\mathbf{C}^\top\,\|\|_F$	0.5	0.004	0.001

Table 1: Population objects

[20]We compare $\mathbf{K}$ to the matrix $\mathbf{A}\mathbf{G}^+$ because our theory describes how we can approximate $\mathbf{K}$ with $\mathbf{\Lambda}\mathbf{\Phi}^+$, and that we can approximate $\mathbf{G}$ with $\mathbf{\Phi}$.

These findings with population objects set the stage for an experiment in which we construct repeated samples of $M \times T$ data matrices $\mathbf{Y}$, to be described in the next subsection.

4.2 Sample Counterparts

We apply our algorithm to samples paths $\{\mathbf{y}_t\}_{t=1}^{T+1}$ generated by state-space system (36). For each sample $j = 1, \ldots J$, we create data matrices $\mathbf{Y}^{(j)}$ and $\mathbf{Y}'^{(j)}$. We apply Pseudo-code 1 to acquire the following objects for sample j: $\mathbf{\Lambda}^{(j)}, \mathbf{\Phi}^{(j)}, \widehat{\mathbf{K}}^{(j)}, \widehat{\mathbf{\Omega}}^{(j)}, \widehat{\mathbf{\Sigma}}_{\infty}^{(j)}, \widehat{\mathbf{R}}^{(j)}, \widehat{\mathbf{C}\,\mathbf{C}}^{\top(j)}$. Then we calculate element-wise means across all samples. For example, we compute

$$\widehat{E}[\mathbf{\Lambda}] \equiv \frac{1}{J} \sum_{j=1}^{J} \mathbf{\Lambda}^{(j)}$$

where $\widehat{E}[\mathbf{\Lambda}]$ is an $N \times N$ matrix of sample means. We then subtract the sample mean from its population counterpart to approximate mean errors of our sampled DMD estimators.

We split the finite-sample comparisons into two distinct, but related parts. The first of these implements the finite-sample analogue of the population exercise in Section 4.1 by fixing $T = 150$ and increasing M. In both exercises, $J = 5000$. Table 3 reports Frobenius norms of mean errors. Generally speaking, the simulations show that the mean errors fall as M increases from 300 to 1000. However this result is more stark in some parameters than others. For example, increasing M reduces the bias in the DMD estimator $\widehat{\mathbf{B}}$ by an order of magnitude, while the reduction in the bias in $\mathbf{\Lambda}$ is not as profound.

Next, we fix $M = 1000$ and increase T from 150 to 999. Table 3 reports outcomes. The simulations show that the mean error falls for all the estimators, especially those for which the increase in M from 300 to 1000 has little effect, such as $\mathbf{\Lambda}$. Taken together,

Object	$M = 300$	$M = 1000$
$\|\|\widehat{E}[\mathbf{\Lambda}] - \mathbf{A}\|\|_F$	0.046	0.043
$M^{-1}\|\|\widehat{E}[\mathbf{\Phi}] - \mathbf{G}\|\|_F$	$4.0e^{-3}$	$2.2e^{-3}$
$M^{-1}\|\|\widehat{E}[\widehat{\mathbf{B}} - \mathbf{B}\|\|_F$	1.7^{-4}	$5.1e^{-5}$
$M^{-1}\|\|\widehat{E}[\widehat{\mathbf{K}}] - \mathbf{K}\|\|_F$	$4.1e^{-5}$	$5.7e^{-6}$
$M^{-1}\|\|\widehat{E}[\mathbf{\Phi}^+] - \mathbf{L}\|\|_F$	$6.4e^{-5}$	$9.3e^{-6}$
$M^{-1}\|\|\widehat{E}[\widehat{\mathbf{\Omega}}] - \mathbf{\Omega}\|\|_F$	5.9^{-3}	$5.7e^{-3}$
$\|\|\widehat{E}[\widehat{\mathbf{\Sigma}}_\infty] - \mathbf{\Sigma}_\infty\|\|_F$	0.92	0.59
$M^{-1}\|\|\widehat{E}[\widehat{\mathbf{R}}] - \mathbf{R}\|\|_F$	0.006	0.005
$\|\|E[\widehat{\mathbf{C}\,\mathbf{C}}^\top] - \mathbf{C}\,\mathbf{C}^\top\|\|_F$	0.94	0.62

Table 2: Estimate of finite-sample bias of DMD estimation for $M = 300$ and $M = 1000$, for $T = 150$

our lab experiments suggest the important role that M and T both play in shaping the quality of inferences based on algorithm 1.

Object	$T = 150$	$T = 999$
$\|\|\widehat{E}[\mathbf{\Lambda}] - \mathbf{A}\|\|_F$	0.043	$6.7e^{-3}$
$M^{-1}\|\|\widehat{E}[\mathbf{\Phi}] - \mathbf{G}\|\|_F$	$2.2e^{-3}$	$8.4e^{-4}$
$M^{-1}\|\|\widehat{E}[\widehat{\mathbf{B}} - \mathbf{B}\|\|_F$	$5.1e^{-5}$	$6.8e^{-6}$
$M^{-1}\|\|\widehat{E}[\widehat{\mathbf{K}}] - \mathbf{K}\|\|_F$	$5.7e^{-6}$	$1.8e-06$
$M^{-1}\|\|\widehat{E}[\mathbf{\Phi}^+] - \mathbf{L}\|\|_F$	$9.3e^{-6}$	$2.4e-06$
$M^{-1}\|\|\widehat{E}[\widehat{\mathbf{\Omega}}] - \mathbf{\Omega}\|\|_F$	$5.7e^{-3}$	$1.0e^{-3}$
$\|\|\widehat{E}[\widehat{\mathbf{\Sigma}}_\infty] - \mathbf{\Sigma}_\infty\|\|_F$	0.59	0.075
$M^{-1}\|\|\widehat{E}[\widehat{\mathbf{R}}] - \mathbf{R}\|\|_F$	0.005	0.001
$\|\|E[\widehat{\mathbf{C}\,\mathbf{C}}^\top] - \mathbf{C}\,\mathbf{C}^\top\|\|_F$	0.62	0.077

Table 3: Estimate of finite-sample bias of DMD estimation for $M = 1000$, for $T = 150$ and $T = 999$

5 US income and consumption dynamics

We construct representation (10)-(11) for quantiles of cross-sections of US consumption and income and use it to describe salient features of US business cycles from 1990 to 2021. We want to study connections between cross-section dynamics and the cross section averages that macroeconomic models are designed to describe and understand.

We characterize cross-section distributions by their quantiles.[21]

5.1 Data matrix Y

We gather data on private income, post-tax income and consumption from quarterly waves of the Interview section of the Consumer Expenditure Survey between 1990 and 2021.

For **Consumption** we sum total expenditure in the current month (TOTEXPCQ) and total expenditure in the previous month (TOTEXPPQ). We do this because, while households are interviewed once every three months, timings may not match calendar quarters. Summing both measures is a common approach that the BLS has suggested.

Private income consists of a subset of categories that sum to before-tax income. Due to changes in definitions over time, the relevant codes are FINCBTAX (1990-2004) and FINCBTXM (2004-2021). FINCBTXM also changed its definition in 2013, but retained the same code name. We categorize the sub-categories of FINCBT* series into either private income or transfer income. Private income is defined as the sum of wages, business income, financial income, income from rental properties and pensions or annuities from any source. Transfer income is all other sub-categories of before-tax income that are not private income. Table 8 in Appendix C describes the categorization.

Post-tax income is private income plus transfers (i.e. before-tax income) minus taxes paid. Relevant codes are FINCATAX (1990-2004), FINCATXM (2004 - 2013) and FINCATXEM (2013-2021).

The Interview survey asks about income over the past 12 months, while it asks about quarterly amounts of consumption. As recommended by the BLS, we divide all income data by four. We remove all households that consume more than their reported annual incomes in a single quarter.

[21] Let $F_t[0, B] \to [0, 1]$ be a cumulative distribution function for a possibly bounded nonnegative random variable at t. Associated with c.d.f. F is a "quantile function" $Q : [0, 1] \to [0, B]$ that under some regularity conditions is the inverse of F. Here we'll work with "percentiles" defined as $Q(\frac{1}{i}), i = 1, \ldots, 100$.

From these data, we constructed time series of cross-section distributions for consumption, private income, and post-tax income. We'll describe how we did this for consumption and note that we followed the same steps to construct panels for the other two variables. For wave t, we rank all $\tilde{I}$ households by their consumption levels $c_{1,t}, c_{2,t}, \ldots, c_{\tilde{I},t}$. Then we split them into one hundred equally sized bins that we call percentiles. If $\tilde{I}$ is divisible by 100, then each percentile contains $I = \frac{\tilde{I}}{100}$ households. If not, the bottom 99 bins each contain I households and the 100th bin contains the remainder. Except for this detail, define $\tilde{q}_{p,t}$ as mean consumption of households in percentile p:

$$\tilde{q}_{p,t} = \frac{1}{I} \sum_{j=1}^{I} c_{j+I(p-1),t} \quad \text{for } p = 1, \ldots, 100$$

Quantity $\tilde{q}_{p,t}$ is the consumption level of the p-th percentile of the consumption distribution. We repeated this step for every wave to obtain a time-series of consumption at each percentile. We noticed suspicious "jumps" in the constructed time-series for private and post-tax income in Q1-2013 and suspect that these come from the aforementioned changes in variable definitions. We responded to this situation by splicing the pre-jump and post-jump data in Q1-2013. We set Q1-2013 consumption equal to Q4-2012 and recursively accumulated subsequent changes in consumption in the post-jump data to the adjusted series. We deflated each time-series by the Personal Consumption Expenditures price index to obtain real objects and then seasonally adjusted. Call the "cleaned", real, seasonally adjusted consumption percentiles $q_{p,t}$.

We removed 1st and 2nd percentile bins, because consumption for those are negative in some periods. Since earnings in the CEX are top coded, we also removed the 99th percentile bins. We formed quarterly growth rates so that our consumption growth variable is

$$y_{p,t}^{cons} = \log q_{p,t} - \log q_{p,t-1} \quad p = 3, \ldots, 98 \text{ and } t = 1, \ldots, T$$

Proceeding in this way for the two income concepts, our final data set is a quarterly, seasonally adjusted time-series of private income, post-tax income, and consumption growth distributions from 1990 to 2021.

We stack our three cross-sections to form matrix $\mathbf{Y}$. Consequently, $M = 291$ and

$$\mathbf{Y} = \begin{bmatrix} y_{3,1}^{private} & \cdots & y_{3,T}^{private} \\ \vdots & \cdots & \vdots \\ y_{98,1}^{private} & \cdots & y_{98,T}^{private} \\ y_{3,1}^{post-tax} & \cdots & y_{3,T}^{post-tax} \\ \vdots & \cdots & \vdots \\ y_{98,1}^{post-tax} & \cdots & y_{98,T}^{post-tax} \\ y_{3,1}^{cons} & \cdots & y_{3,T}^{cons} \\ \vdots & \cdots & \vdots \\ y_{98,1}^{cons} & \cdots & y_{98,T}^{cons} \end{bmatrix} \tag{37}$$

To help us to interpret the "dynamic modes" that we shall recover from our reduced-rank first-order VAR, we calculate seasonally adjusted growth rates of the first two moments of the private income, post-tax income and consumption distributions. We describe how we do this for the consumption series, again noting that we used the same procedure to compute cross section moments for the two income concept series.

To obtain one cross-section moment, we take the cross-section mean of consumption $\bar{q}_t := \frac{1}{97}\sum_{p=3}^{98} q_{p,t}$. Then we seasonally adjust and take log differences $\log \bar{q}_t - \log \bar{q}_{t-1} =: \mu_t^{cons}$. (Taking means across quantiles $q_{p,t}$ or households $c_{i,t}$ yields the same quantities.)

To obtain another moment, we compute the cross-section variance of consumption $s_t := \frac{1}{97}\sum_{p=3}^{98}(q_{p,t} - \bar{q}_t)^2$. Then we seasonally adjust and take log differences $\log s_t - \log s_{t-1} =: \sigma_t^{cons}$.[22] The final objects, μ_t^{cons} and σ_t^{cons}, represent growth rates of the first

[22] An alternative procedure would compute the variance across households $s_t := \frac{1}{I}\sum_{i=1}^{I}(c_{i,t} - \bar{q}_t)^2$. Note that in constructing σ_t^{cons}, we are ignoring the variances within each percentile bin. We also compute the variance from percentiles as described in Chang et al. (2021, Appendix A) with very similar empirical results.

two moments of the consumption distribution.

5.2 Reduced-Rank First-Order VAR

We want to focus on two dominant modes, so we set $N = 2$, apply the steps in Section 2 to data matrix $\mathbf{Y}$, and extract what DMD researchers call "dynamic modes", i.e., the N components of $\widetilde{\mathbf{x}}_t = \mathbf{\Phi}^+ \mathbf{y}_t$. Thus, mode i is $\widetilde{\mathbf{x}}_{i,t} = \mathbf{\Phi}^+_{:,i} \mathbf{y}_t$, where $\mathbf{\Phi}^+_{:,i}$ is the ith column of the $M \times N$ matrix $\mathbf{\Phi}^+$. The modes evolve according to equations (10) or of the corresponding dynamic system with one-step ahead prediction errors in the dynamic modes having been orthogonalized via a Cholesky decomposition[23], i.e.

$$\widetilde{\mathbf{x}}_{t+1} := \mathbf{\Lambda} \widetilde{\mathbf{x}}_t + \mathbf{H}\, \mathbf{e}_{t+1} \tag{38}$$

where $\mathbf{e}_t$ is a standardized random vector. Our algorithm infers

$$\mathbf{\Lambda} = \begin{bmatrix} 0.83 & 0 \\ 0 & 0.72 \end{bmatrix}, \quad \mathbf{H}\,\mathbf{H}^\top = \begin{bmatrix} 0.11 & 0.05 \\ 0.05 & 0.13 \end{bmatrix}, \tag{39}$$

so both modes are persistent and their innovations have a similar variances. Since $\mathbf{H}\,\mathbf{H}^\top$ is not diagonal, the modes are correlated. The one-step ahead conditional correlation between the two modes is 0.48. By setting $h = 100$ in formula (12) we approximate an unconditional covariance matrix to be

$$\begin{bmatrix} 0.33 & 0.14 \\ 0.14 & 0.26 \end{bmatrix}$$

with an associated unconditional correlation of 0.46 between the two modes. These objects will be useful below when we try to reconcile our findings with ones discovered

[23] Therefore, we define $\mathbf{H}\,\mathbf{H}^\top = \mathbf{\Phi}^+ \mathbf{\Omega} (\mathbf{\Phi}^+)^\top$. This definition and its interpretation is further reinforced in Section 3.

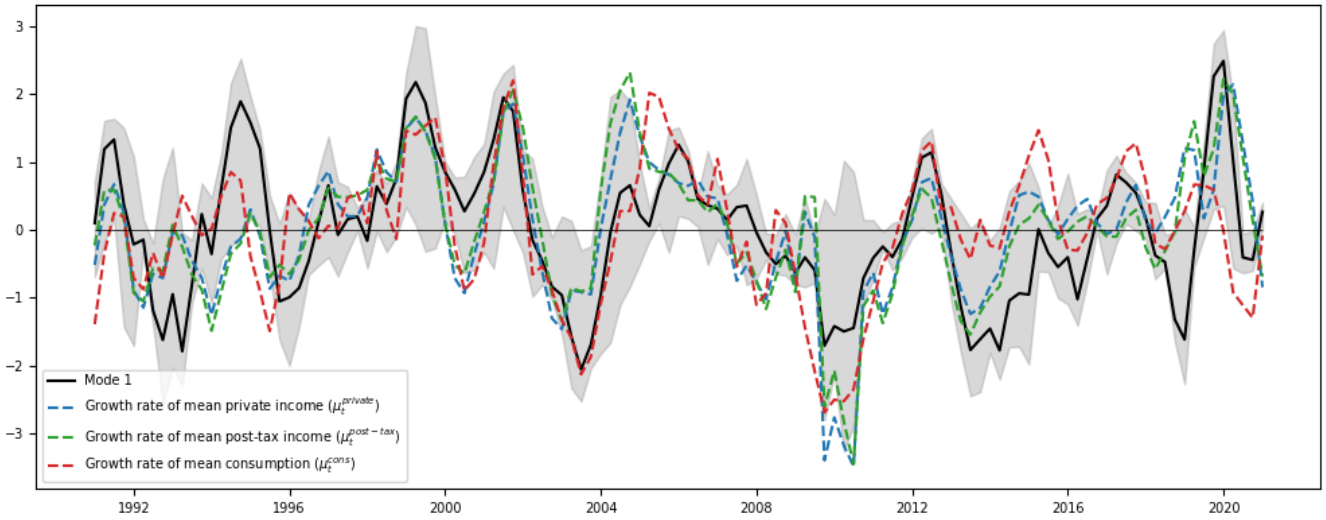

Figure 1: Mode 1, $\tilde{\mathbf{x}}_{1,t} = \mathbf{\Phi}^{+}_{:,1} y_t$ (standardized)

,

with other approaches.

Figure 1 displays a mode $\widetilde{\mathbf{x}}_{1,t}$, standardized to zero mean and unit standard deviation. The dark line in Figure 1 plots mode $\widetilde{\mathbf{x}}_{1,t}$, i.e., the inner product of $\mathbf{y}_t$ with $\mathbf{\Phi}^{+}_{:,1}$; the three colored dashed lines show standardized growth of average private income ($\mu_t^{private}$), after-tax income ($\mu_t^{post-tax}$) and consumption (μ_t^{cons}) defined at the bottom of section 5.1. Gray denotes a 95% confidence interval of mode 1, computed using the recursive bootstrap with 5000 simulation draws (see Hansen (2022, sec. 14.46)).[24]

Table 4 shows the results of six bivariate regressions of the growth of average incomes and consumption (the dashed lines in Figure 1) on modes 1 and 2. The coefficients on both modes are significant and positive in all regressions, and the R^2 is also substantial. The R^2 associated with mode 1 is higher than mode 2 for consumption, and vice versa for income. Evidently, mode 1 is highly correlated with cross-section averages for all three cross-sections, so it seems to qualify as a candidate for a Burns

[24]We simulate bootstrap samples of $\mathbf{y}_t^b$ according to $\mathbf{y}_{t+1}^b = \widehat{\mathbf{B}}\,\mathbf{y}_t^b + \mathbf{a}_{t+1}^b$ for $t = 1, \ldots, T$, where $\mathbf{a}_{t+1}^b$ is drawn with replacement from recentered residuals in (8), $\mathbf{a}_{t+1}^* = \widehat{\mathbf{a}}_{t+1} - \frac{1}{T}\sum_{t=1}^{T} \widehat{\mathbf{a}}_{t+1}$. With this bootstrap sample, we execute steps 1.-3. in Section 2 to obtain $\mathbf{\Phi}^b$. Then we compute the bootstrap modes $\widetilde{\mathbf{x}}_{t+1}^b = (\mathbf{\Phi}^b)^{+}\,\mathbf{y}_t$. We repeat this 1000 times and construct an associated distribution of $\widetilde{\mathbf{x}}_{t+1}^b$. We report the 95% coverage intervals for that distribution.

	$\mu_t^{private}$		$\mu_t^{post-tax}$		μ_t^{cons}	
Constant	0.32	0.32	0.36	0.36	0.19	0.19
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Mode 1	0.66		0.64		0.49	
	(0.08)		(0.07)		(0.07)	
Mode 2		0.85		0.81		0.42
		(0.07)		(0.06)		(0.07)
R-squared	0.37	0.55	0.40	0.57	0.32	0.22

Table 4: Regression results of growth of first moment of distributions (μ_t) on modes

and Mitchell's unidimensional "reference cycle" index for business cycles.

To indicate how quantiles of our three cross-sections respond to mode 1, figure 2 plots the first column of $\mathbf{\Phi}$. All components (or "loadings") are positive: as mode 1 rises, so do all quantiles of incomes and consumption.

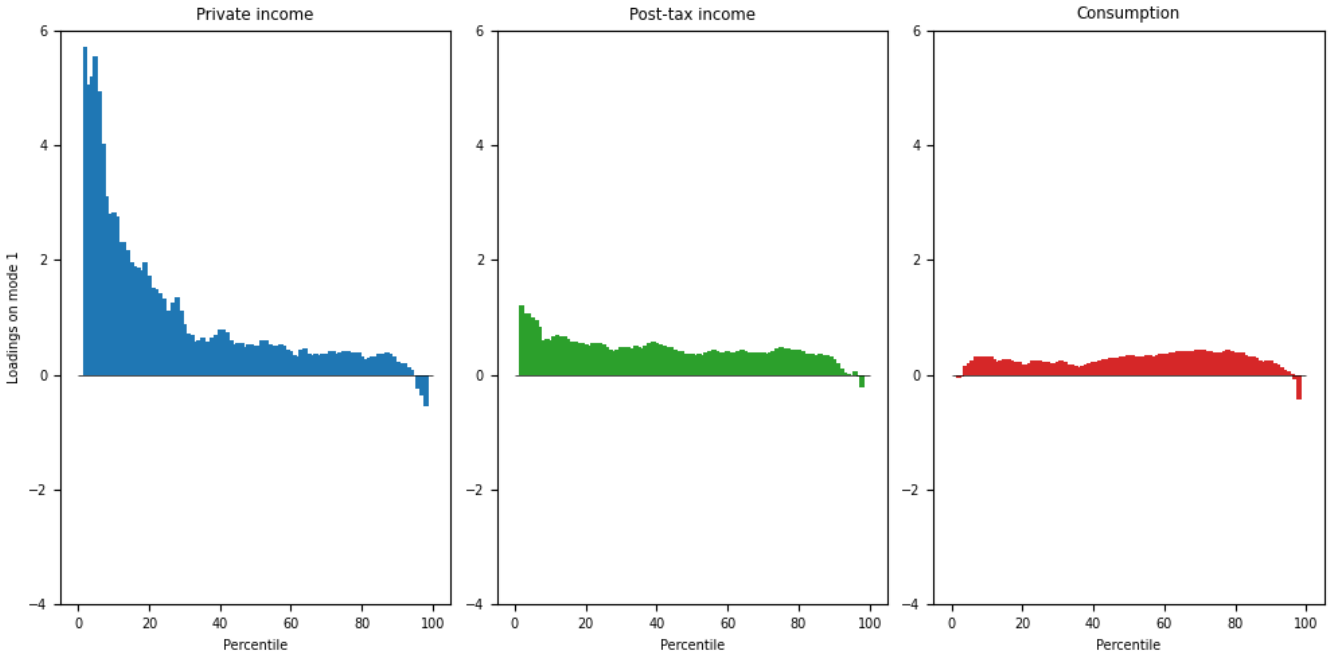

Figure 2: Loadings $\mathbf{\Phi}_{\cdot,1}$ of quantiles on mode 1

Low quantiles of private income growth are more sensitive to mode 1 than are high quantiles. All quantiles of total income growth are less sensitive to mode 1 than are corresponding quantiles of private income growth. Most consumption quantiles are

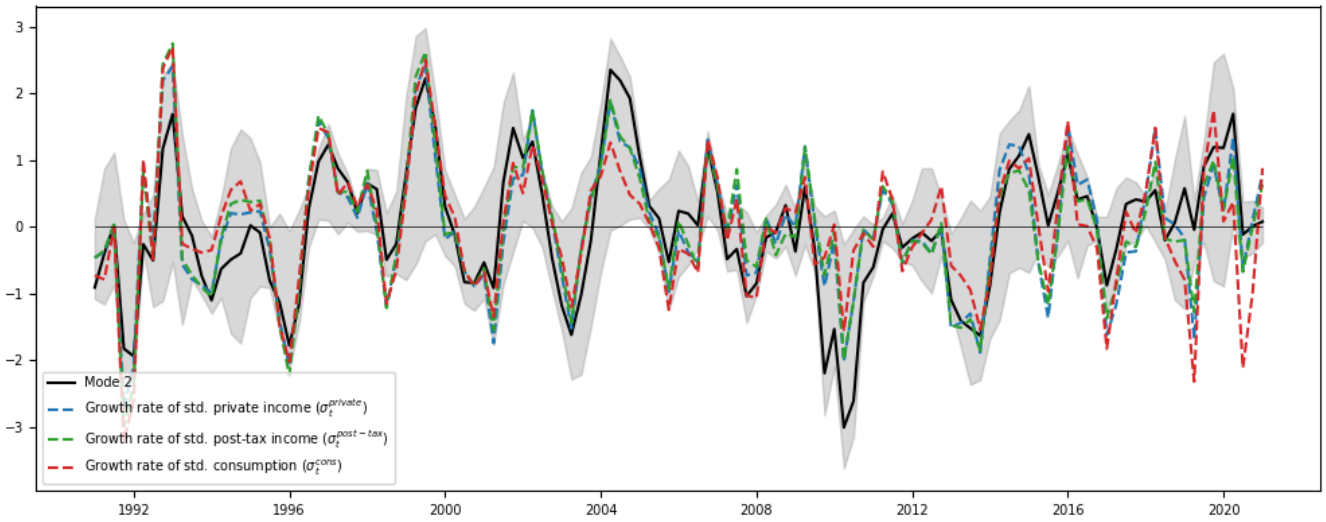

Figure 3: Mode 2, $\tilde{\mathbf{x}}_{2,t} = \mathbf{\Phi}^{+}_{:,2} y_t$ (standardized)

even less sensitive.

Figure 3 shows mode $\tilde{\mathbf{x}}_{2,t}$. The dark line in Figure 3 plots standardized mode $\widetilde{\mathbf{x}}_{2,t}$ while the three colored dotted lines show standardized growth rates of the cross-section standard deviations of income and consumption variables.[25] Table 5 shows counterparts to the Table 4 regressions that use cross-section standard deviations, instead of cross-section averages. Coefficients on both modes are positive and significant, and R^2 statistics of the mode 2 regressions are high for all three variables.

	$\sigma_t^{private}$		$\sigma_t^{post-tax}$		σ_t^{cons}	
Constant	0.39	0.39	0.43	0.42	0.11	0.10
	(0.11)	(0.06)	(0.12)	(0.06)	(0.13)	(0.09)
Mode 1	0.50		0.76		0.70	
	(0.20)		(0.20)		(0.21)	
Mode 2		2.1		2.19		1.98
		(0.10)		(0.11)		(0.15)
R-squared	0.05	0.77	0.11	0.79	0.08	0.59

Table 5: Regressions of growth of second moment of distributions (σ_t) on modes

To indicate how quantiles of our three cross-sections respond to mode 2, figure

[25] As before, the gray swathe denotes the 95% confidence interval of mode 2, computed using the recursive bootstrap Hansen (2022, sec. 14.46) with 5000 simulation draws.

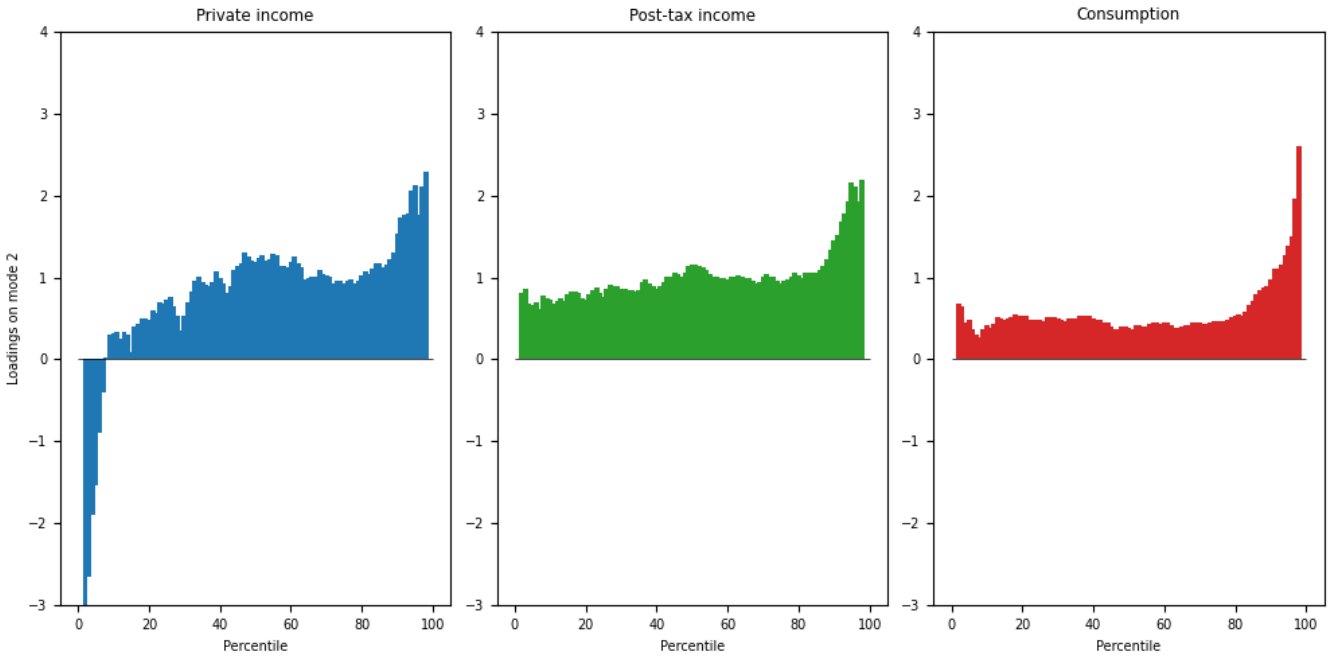

Figure 4: Loadings $\mathbf{\Phi}_{\cdot,2}$ of quantiles on mode 2

2 plots the second column of $\mathbf{\Phi}$. When mode 2 rises, higher quantiles of all three cross-sections increase more than lower quantiles. Private income growth falls for low quantiles. Mode 2 seems to be an "inequality mode".

5.3 Connections with other data summaries

Our DMD detects disparate responses of income and consumption quantiles to two dynamic modes. The reference cycle mode 1 affects low quantiles of private income the most, and provokes smaller responses of other quantiles. In contrast, responses of all consumption quantiles to model 1 are similar.

Several papers have inferred that higher quantiles consumption are more responsive to business cycles than lower ones. Parker and Vissing-Jorgensen (2009) regress the annual change in log consumption on the coincident change in log aggregate real per capital consumption using CEX data. They find that the sensitivity of the top 10% of the distribution is substantially higher than of the lower 80%. They find a qualitatively

similar relationship for income using data from Piketty and Saez (2003). Guvenan et al. (2014) and Guvenan et al. (2017) compute a similar regression on Social Security Administration data and find that exposures of earnings to aggregate variables are "U-shaped" with respect to the earnings level.

Are our results compatible with these? To find out, we substitute our DMD estimates of $\mathbf{\Lambda}, \mathbf{H},$ and $\mathbf{\Phi}$ in system (38) and (11) and simulate data of length $T = 1000$. For every period t and each variable, we calculate cross-sectional means, i.e.,

$$\bar{y}_t^{private} = \frac{1}{97}\sum_{i=3}^{98} y_{i,t}^{private}, \quad \bar{y}_t^{post-tax} = \frac{1}{97}\sum_{i=3}^{98} y_{i,t}^{post-tax}, \quad \bar{y}_t^{cons} = \frac{1}{97}\sum_{i=3}^{98} y_{i,t}^{cons}$$

We then compute time-series regressions of percentiles on the corresponding aggregate. For private income, the regression is

$$y_{i,t}^{private} = \alpha_i^{private} + \beta_i^{private}\bar{y}_t^{private} + \epsilon_{i,t}^{private} \quad \text{for } i = 3, \ldots, 98 \tag{40}$$

Figure (5) plots β_i as a function of the percentile rank.[26] The left chart is for private income and shows that households in the low quantiles have high betas on aggregate private income growth; those in the middle of the distribution have lower betas, while betas are slightly higher for upper quantiles. This pattern is qualitatively similar to the "U-shaped" betas described by Guvenan et al. (2014).[27] The right panel shows corresponding betas for the consumption distribution. It shows that the highest quantiles of consumption are most responsive to aggregate consumption growth, in line with the findings of Parker and Vissing-Jorgensen (2009) and others. Thus, computing regressions like those in the literature on data simulated from our estimated DMD model yields similar results.

[26] Almost identical plots arise when we weight quintile growth rates by the corresponding income/consumption shares. We compute simple means in favor of clarity.

[27] Since the Social Security Administration data used by Guvenan et al. (2014) is not top coded, the authors are able to include very high income individuals – the top 0.1% of households – in their analysis. This difference might explain why our income betas are not as distinctly "U-shaped".

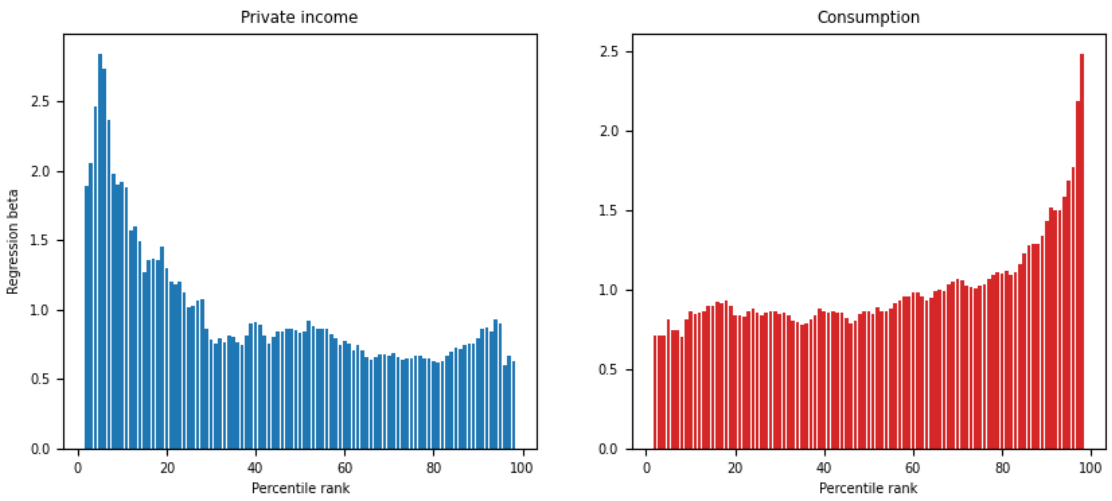

Figure 5: Aggregate regression betas

According to our DMD, the aggregate variables used in regressions (40) covary along with two correlated dynamic modes. Consequently, estimated β_is are functions of DMD loadings $\mathbf{\Phi}_{[i,:]}$ on both modes. Increasing consumption loadings on inequality mode 2 drive the increasing β_i^{cons}, not loadings on the reference cycle mode 1. A similar effect, albeit smaller quantitatively appears in the left panel. Our DMD model indicates that those "U-shaped" private income betas found by Guvenan et al. (2014) are driven by fluctuations in the inequality mode 2, not the reference cycle mode 1.

5.4 Impulse responses

To construct impulse response functions with orthogonal shocks, we again adopt representation (38) - (11). Figure 6 plots impulse responses to a 1 standard deviation increase in the first orthogonalized shock. Since innovations $\widehat{\mathbf{a}}_t$ to the modes are correlated, the first shock affects both modes (top left panel). The first shock affects the lower private income quantiles much more than it does the higher quantiles. Differences in responses are much lower for after-tax income and consumption quantiles.

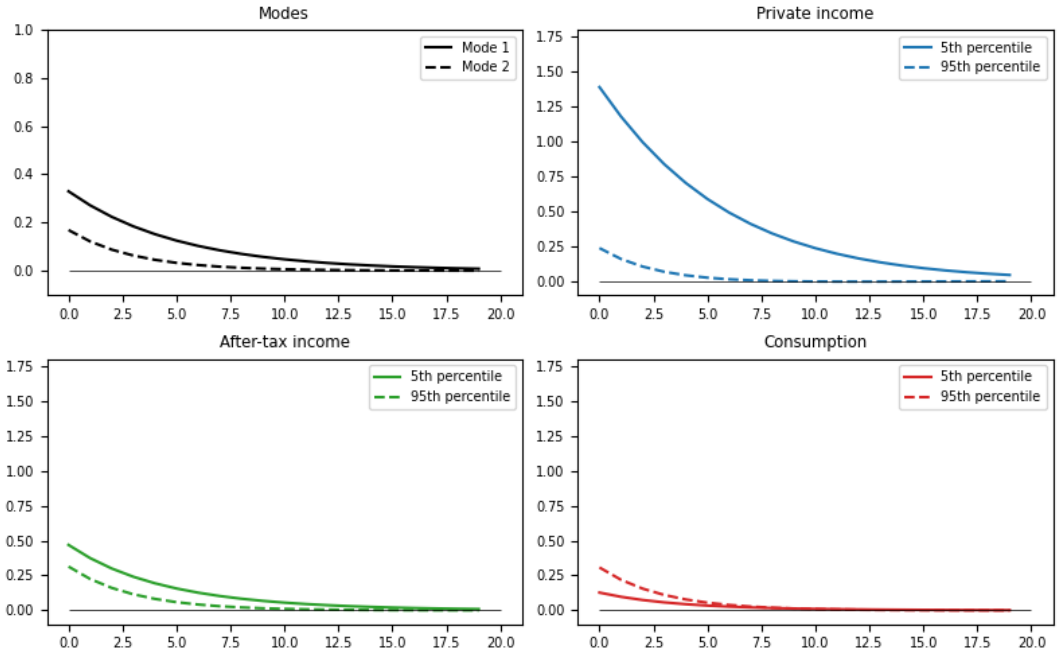

Figure 6: Impulse response to orthogonalized shock 1

The impulse responses also imply differential effects to income and consumption inequality. Figure 7 plots the impulse response of the difference between the 95th percentile and the 5th percentile to a one-standard deviation **decrease** in shock 1. It shows that private income growth inequality increases substantially, and consumption inequality falls. These results are also consistent with findings contained in papers that study business cycle dynamics of inequality. Using CPS data, Meyer and Sullivan (2013) find income inequality rose during the 2009 recession, while consumption inequality fell. In a more recent paper, Meyer and Sullivan (2023) confirm their earlier findings and attribute this dichotomy to asset ownership, the prices of which fell dramatically during the 2009 recession. In response to the negative shock, post-tax income inequality rises, albeit much smaller in magnitude than private income. Its reaction is consistent with government transfers being an important redistribution mechanism, as documented in Heathcote et al. (2023).[28]

[28] Heathcote et al. (2023) also infer the importance of household income pooling in the muting of consumption

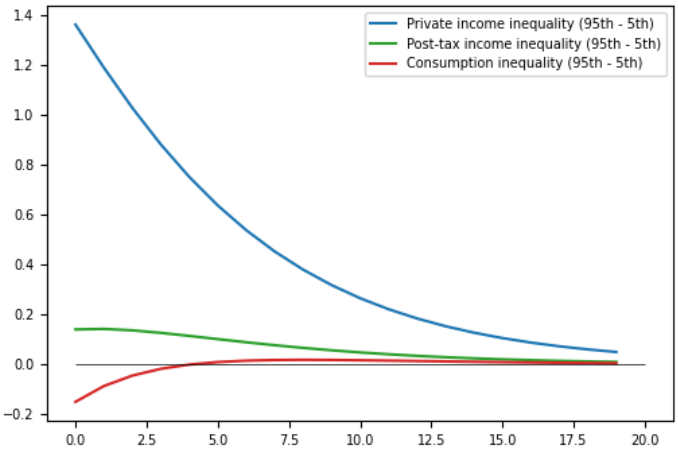

Figure 7: Impulse response of 95th - 5th percentile income and consumption to *decrease* in shock 1

Another implication of the impulse response to shock 1 is that consumption inequality – defined as the cross-sectional standard deviation – arising from shock 1 is "smoother" than income inequality. To quantify this, we simulate a long sample ($T = 1000$) from system (38)-(11), setting shock 2 always equal to zero. For each period, we calculated an inequality measure for each variable, defined as a cross-sectional standard deviation. The result is a time-series of our measure of inequality $\{\hat{\sigma}_t\}$ for each variable. Table 6 shows sample variances of inequality for all three variables for the simulated data; these are our estimates of the variances of our inequality measures attributable to shock 1. [29]

The variance of private income inequality is higher than are the variances of post-

inequality. Since our private income variable is already at the household level, our current analysis is unable to speak to that insight.

[29] Explicitly, for consumption, the calculation is

$$\frac{1}{1000}\sum_{t=1}^{1000}\left(\hat{\sigma}_t^{cons} - \frac{1}{1000}\sum_{t=1}^{1000}\hat{\sigma}_t^{cons}\right)^2$$

The analagous computations are also done for private income and post-tax income.

	Private income	Post-tax income	Consumption
Variance of inequality	0.11	0.002	0.001

Table 6: Estimated variance of inequality due to shock 1 from simulated data

tax income and consumption. This is consistent with findings of Heathcote et al. (2010) (see figure 13), who plot the different measures of inequality in disposable income and non-durable consumption between 1980 and 2005. These inequality measures include the variance of the log, the Gini coefficient, the P50-P10 ratio and the P90-P50 ratio. In all four cases, the time-series of consumption inequality appears, to the naked eye, less volatile than that of disposable income.

Figure 8 plots impulse responses to a 1 standard deviation increase of the second orthogonalized shock. By virtue of the lower triangular cholesky decomposition, the second shock affects only the second mode. Across our three variables, differences in responses of the 5th and 95th percentiles are largest for private income. For private income, the 5th quantile falls by 0.6ppts while the 95th quantile *increases* by a similar magnitude. It seems that shock 2 "redistributes" private income. Not so for post-tax income and consumption quantiles. Responses of all quantiles are positive, though they are larger in magnitude for higher quantiles.

5.5 State-space objects recovered from reduced-rank first-order VAR

As presented in Section 3, estimates of $\mathbf{A}$ and $\mathbf{G}$ come directly from the DMD estimates in (39) and Figures 2 and 4. We use steps 10 and 12 of Pseudo-code 1 to recover shock covariance matrix $\mathbf{C}\,\mathbf{C}^\top$ and conditional covariance $\mathbf{\Sigma}_\infty$. We use step 11 of Pseudo-code 1 to estimate $\mathbf{R}$. We obtain

$$\widehat{\mathbf{C}\,\mathbf{C}}^\top = \begin{bmatrix} 0.018 & 0.015 \\ 0.015 & 0.055 \end{bmatrix} \quad \widehat{\mathbf{\Sigma}}_\infty = \begin{bmatrix} 0.055 & 0.030 \\ 0.030 & 0.078 \end{bmatrix},$$

while Figure 9 plots the diagonal of $\widehat{\mathbf{R}}$, the variance of the measurement error. Measure-

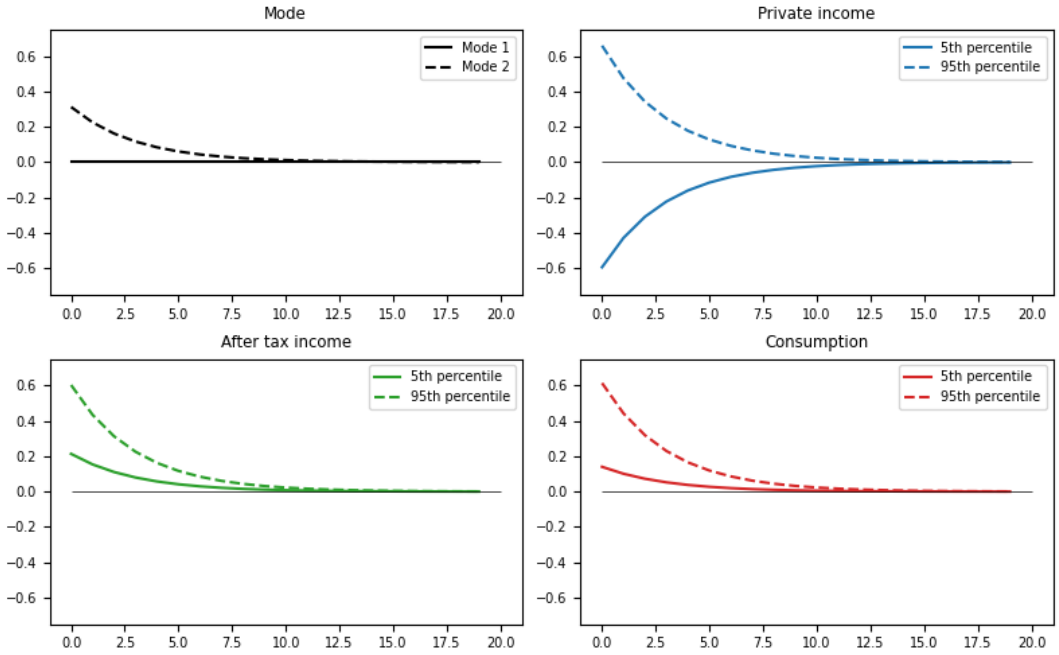

Figure 8: Impulse response to orthogonalized shock 2

ment error variances are highest at the very low and very high quantiles, and relatively small for middle quantiles.

We use equations (34) and (35) to compute $\mathbf{V_y}$. We decompose unconditional variances diag($\mathbf{V_y}$) into parts attributable to the factors (diag($\mathbf{G}\,\mathbf{V_x}\,\mathbf{G}^\top$)) and parts attributable to the measurement error diag($\mathbf{R}$). We present these in Figure 10. In the middle quantiles for both income variables, the proportion of the variance explained by the factors is around 50%. This proportion falls to around 20% for both low and high quantiles. For consumption, for most quantiles only around 20% of the variance is explained by the factors.

To form a frequency-by-frequency counterpart of variance decomposition (35), we

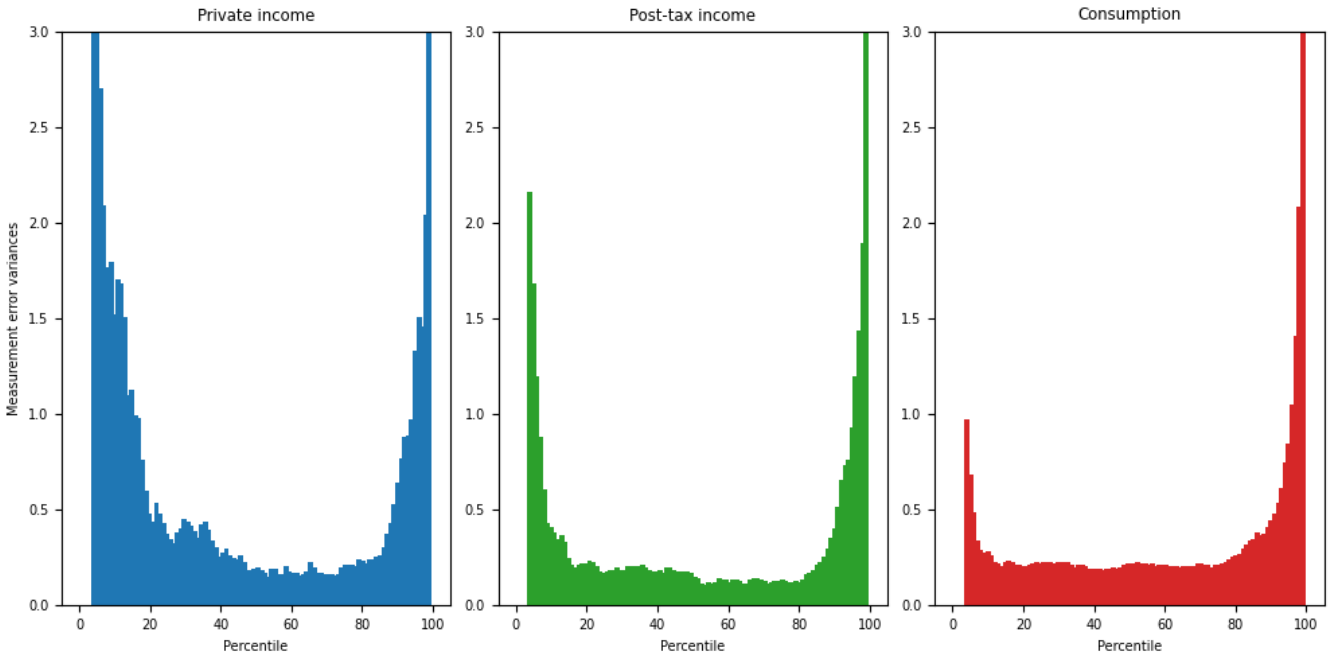

Figure 9: Measurement error variances

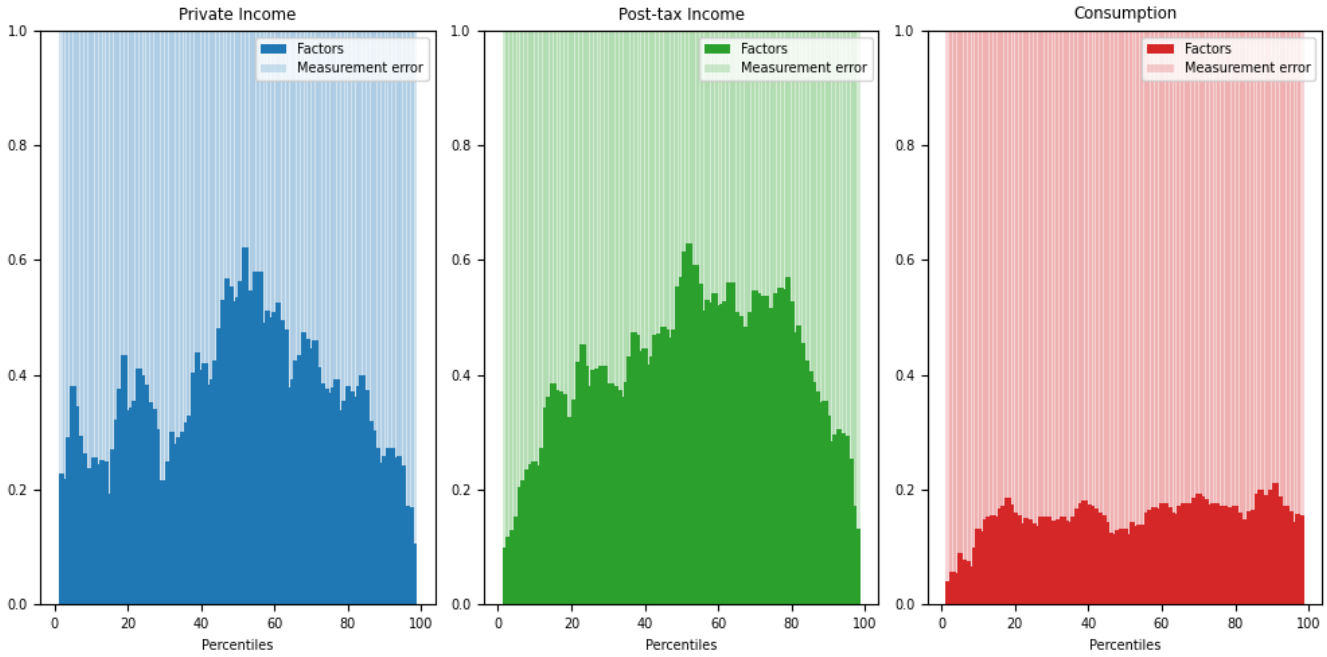

Figure 10: Decomposition of unconditional variance $\mathbf{V_y} = \mathbf{G}\,\mathbf{V_x}\,\mathbf{G}^\top + \mathbf{R}$ of $\mathbf{y}$.

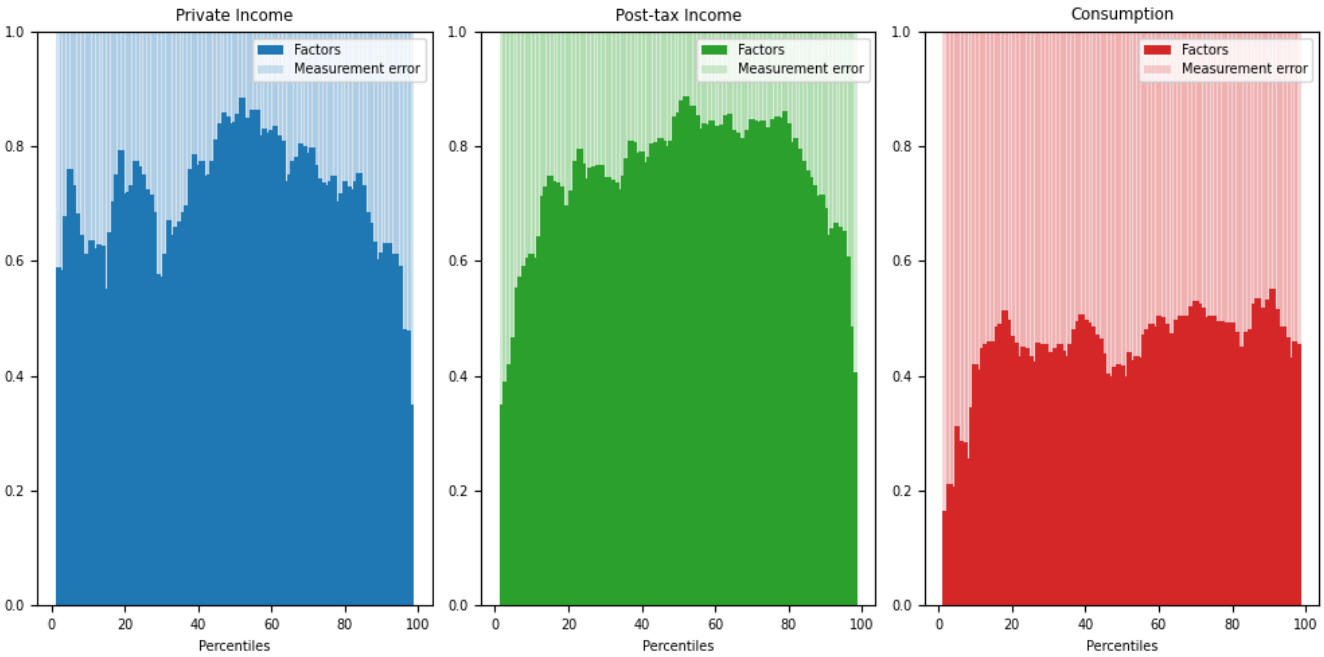

Figure 11: Spectral density decomposition of $\mathbf{y}$ at eight-year frequency ($\omega = \frac{2\pi}{32}$)

form spectral densities of processes $\mathbf{x}$ and $\mathbf{y}$ at at frequencies $\omega \in [0, 2\pi]$:

$$S_{\mathbf{x}}(\omega) = [\mathbf{I} - \mathbf{A}\, e^{-\omega j}]^{-1}\, \mathbf{C}\, \mathbf{C}^{\top} [\mathbf{I} - \mathbf{A}^{\top}\, e^{\omega j}]^{-1}$$
$$S_{\mathbf{y}}(\omega) = \mathbf{G}\, S_{\mathbf{x}}(\omega)\, \mathbf{G}^{\top} + \mathbf{R} \tag{41}$$

Figures 11 and 12 decompose $\text{diag}(S_{\mathbf{y}}(\omega))$ into a part $\text{diag}(\mathbf{G}\, S_{\mathbf{x}}(\omega)\, \mathbf{G}^{\top})$ attributable to the factors and a part $\text{diag}(\mathbf{R})$ attributable to measurement errors at frequencies corresponding to periods of 8 years and 20 years, respectively.[30] Figures 11 and 12 indicate that the two hidden factors explain large percentages of the variances of quantiles at 20-year and 8-year frequencies.

[30] Since our data is quarterly, for a eight-year period $\omega = \frac{2\pi}{32}$ and for a twenty-year period $\omega = \frac{2\pi}{80}$

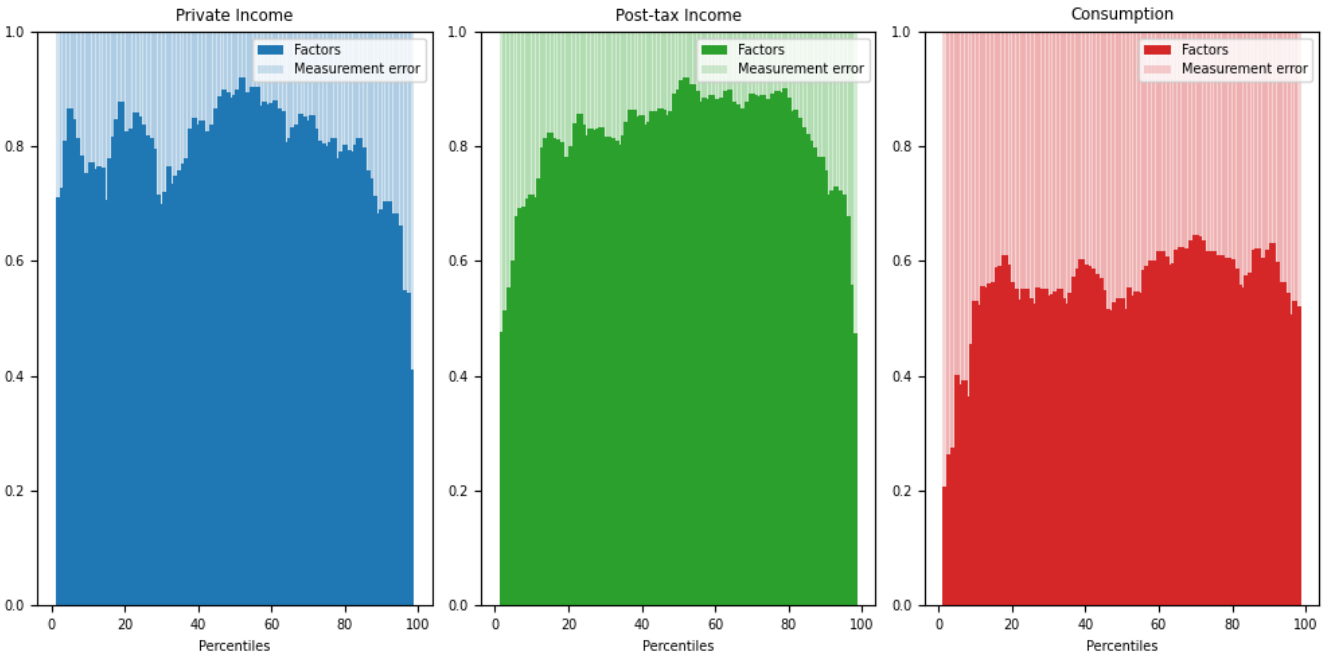

Figure 12: Spectral density decomposition of $\mathbf{y}$ at twenty-year frequency ($\omega = \frac{2\pi}{80}$)

5.6 Technical qualification

In this section we have adhered to a widespread practice described by Stock and Watson (2016, Sec. 2) that forms first differences of all variables in order to validate the assumption that $\{\mathbf{y}_t\}$ is a covariance stationary process. As Stock and Watson note, this practice is vulnerable to the criticism that by over-differencing, it obscures information that could tighten inference about parameters of the dynamic factor model. In a sequel to this work, we shall respond to this criticism by combining the DMD procedures deployed here with the additive functional structures presented by Hansen (2012). For now, we simply say that, in the spirit of Stock and Watson, we find both approaches potentially useful.

6 Concluding remarks

We have described restrictions on parameters of an LQG hidden Markov model that allow us to infer its free parameters from a reduced-rank first-order vector autoregression that can be computed via a Dynamic Mode Decomposition and methods provided by Tu et al. (2014).

We use a DMD to represent dynamics of income and consumption cross sections from the Consumer Expenditure Survey. We detect two dominant, correlated dynamic modes: one seems to be a Burns and Mitchell (1946) reference cycle; the other seems to be an inequality factor. Loadings of quantiles of CEX cross section on these modes indicate substantial government redistribution and consumption insurance. Impulse responses of quantiles to innovations in the first "reference cycle" mode indicate quite different responses in earned income and consumption quantiles, another indication of government redistributions.

Iao and Selvakumar (2024) use techniques we have described here as ingredients of an "indirect inference" strategy in the spirit of Gallant and Tauchen (1996) to estimate structural parameters of a heterogeneous agent dynamic general equilibrium model that Iao and Selvakumar can simulate, but whose likelihood function is too expensive for them to compute. So they approach the data with two models, namely, their structural model and another statistical model whose likelihood they can compute. Thus, their procedure uses two collections (manifolds) of probability distributions for outcomes $x \in X$, namely, a DMD manifold of statistical models $\{f_\theta(x)\}_{\theta \in \Theta}$ and a manifold of structural macro statistical models $\{g_\delta(x)\}_{\delta \in \Delta}$. Iao and Selvakumar's first step is to estimate parameters $\theta \in \Theta$ of their auxiliary model using our section 3, algorithm 1 DMD procedures. Their second step is to use repeated Monte Carlo simulations of model g_δ to find parameters of $\delta \in \Delta$ of their structural macro model

that minimizes Kullback-Leibler divergence

$$\mathrm{KL}(g_{\delta_o}, f_\theta) = \int \log\left(\frac{g_{\delta_o}(x)}{f_\theta(x)}\right) g_{\delta_o}(x)dx.$$

Their application thereby uses observed cross-section dynamics to estimate structural parameters that govern aggregate dynamics.

A Appendix

A.1 Other Applications of PCA

Stock and Watson (2016, Sec. 2) describe how a principal components representation of a 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 algorithm our calculations.[31] We use a reduced SVD $\mathbf{Y} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}$ to form submatrices $\mathbf{U}$ of $\widetilde{\mathbf{U}}$ and $\mathbf{V}$ of $\widetilde{\mathbf{V}}$ that appear in formula (6) that we use to compute a compressed data matrix $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ that we then use as an input to forming $\widehat{\mathbf{B}} = \mathbf{Y}'\,\mathbf{Y}^+$. We use $\mathbf{U}$ and $\mathbf{V}$ again when we form $\widetilde{\mathbf{B}}_{N\times N} = \mathbf{U}^\top\,\widehat{\mathbf{B}}\,\mathbf{U}$. At this point we apply another eigendecomposition

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

where columns of the $N \times N$ matrix $\mathbf{W}$ are eigenvectors of $\widetilde{\mathbf{B}}$ and eigenvalues of $\widetilde{\mathbf{B}}$ 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) 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 $C_{\mathbf{y}}$, somehow selecting 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.

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

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

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 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).

Comparison of identification assumptions PCA estimators of factors and the associated loadings are only identified up 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 the top $N \times N$ block of $\mathbf{G}$ is lower diagonal, and 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 require that either the factors are orthogonal, the loadings are orthogonal or have some special structure that means one or more observations are noisy measurements of the factors.

Our identification restrictions for DMD impose that the factors jointly follow a VAR(1), with a diagonal $\mathbf{A}$ matrix, but a possibly non-diagonal $\mathbf{C}\mathbf{C}^\top$ matrix. Our restrictions do not require that columns of $\mathbf{G}$ are orthogonal nor do they require that one or more observations are noisy measurements of the factors. We summarize these comparisons in Table 7.

Identifying restrictions		
	Principal components	**DMD**
PC Option 1	• $\frac{1}{T}\widetilde{\mathbf{x}}_t'\widetilde{\mathbf{x}}_t = I_N$ • $\mathbf{G}'\mathbf{G}$ is a diagonal matrix with distinct entries	• $\widetilde{\mathbf{x}}_t$ follows a VAR(1) with – $\mathbf{A}$ is diagonal – $\mathbf{C}$ unrestricted • $\mathbf{G}$ full column rank • $\|\|\mathbf{G}^\top\mathbf{G}\|\| = \mathcal{O}(M)$
PC Option 2	• $\frac{1}{T}\widetilde{\mathbf{x}}_t'\widetilde{\mathbf{x}}_t = I_N$ • $\mathbf{G} = \begin{pmatrix}\mathbf{G}_1\\ \mathbf{G}_2\end{pmatrix}$ • $\mathbf{G}_1 = \begin{pmatrix}\mathbf{g}_{11} & 0 & \cdots & 0\\ \mathbf{g}_{21} & \mathbf{g}_{22} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{g}_{r1} & \mathbf{g}_{r2} & \cdots & \mathbf{g}_{rr}\end{pmatrix}, \mathbf{g}_{ii} \neq 0, i = 1,\ldots,r$	
PC Option 3	• $\widetilde{\mathbf{x}}_t$ unrestricted • $\mathbf{G} = \begin{pmatrix}I_N\\ \mathbf{G}_2\end{pmatrix}$	

Table 7: Identifying restrictions for method of PC and DMD

B A computational detail

Because $M > T$, covariance matrix $\widehat{\mathbf{\Omega}}$ of the sample residuals $\widehat{\mathbf{a}}_1, \ldots, \widehat{\mathbf{a}}_T$ is ill-conditioned. This poses a difficulty for the matrix inversion called for in step 10 of algorithm 1. We cope with this situation by computing a generalized inverse via a Singular Value Decomposition (SVD) of $\widehat{\mathbf{\Omega}}$. Recycling previous SVD notation, let

$$\widehat{\mathbf{\Omega}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^\top$$

For a k that we set, we construct $\mathbf{U} = \widetilde{\mathbf{U}}[:, : k]$, $\mathbf{\Sigma} = \widetilde{\mathbf{\Sigma}}[: k, : k]$ and $\mathbf{V}^\top = \widetilde{\mathbf{V}}^\top[: k, :]$. Then $\widehat{\mathbf{\Omega}}^{-1} = \mathbf{V}\,\mathbf{\Sigma}^{-1}\,\mathbf{U}^\top$ where $\mathbf{\Sigma}^{-1} = \text{diag}(\frac{1}{\mathbf{\Sigma}_{1,1}}, \ldots, \frac{1}{\mathbf{\Sigma}_{k,k}})$.

C Data

	Code Mneumonic		
	1990-2004	2004-2013	2013-2022
Private income			
Income from salary or wages	FSALARYX	FSALARYM	FSALARYM
Income from non-farm business	FNONFRMX	FNONFRMM	FSMPFRXM
Income from own farm	FFRMINCX	FFRMINCM	
Income from interest on savings accounts or bonds	INTEARNX	INTEARNM	INTRDVXM
Regular income earned from dividends, royalties, estates	FININCX	FININCXM	ROYESTXM
Income from pensions or annuities	PENSIONX	PENSIONM	RETSURVM
Net income or loss received from roomers or boarders	INCLOSSA	INCLOSAM	
Net income or loss received other rental properties	INCLOSSB	INCLOSBM	NETRENTM
Income from regular contributions from alimony and other	ALIOTHX	ALIOTHXM	
Income from care of foster children, cash scholarships	OTHRINCX	OTHRINCM	OTHRINCM
Transfer income			
Income from Social Security benefits and Railroad Benefit checks	FRRETIRX	FRRETIRM	FRRETIRM
Supplemental Security Income from all sources	FSSIX	FSSIXM	FSSIXM
Income from unemployment compensation	UNEMPLX	UNEMPLXM	
Income from workmen's compensation and veteran's payments	COMPENSX	COMPENSM	OTHREGXM
Income from public assistance including job training	WELFAREX	WELFAREM	WELFAREM
Income from other child support	CHDOTHX	CHDOTHXM	
Food stamps	JFDSTMPA		
Food stamps and electronic benefits	FOODSMPX	FOODSMPM	JFSAMTM

Table 8: Categorizing CEX income into private and transfers

References

ANDERSON, T. W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *The Annals of Mathematical Statistics*, 22, 327–351.

——— (1999): "Asymptotic distribution of the Reduced Rank Regression Estimator Under General Conditions," *The Annals of Statistics*, 27, 1141–1154.

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

BAI, J. AND S. NG (2008): "Large Dimensional Factor Analysis," *foundations and Trends in Econometrics*, 3, 89–163.

——— (2013): "Principle 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 Learnings, Dynamical Systems, and Control, second edition*, Cambridge University Press.

BURNS, A. F. AND W. C. MITCHELL (1946): *Measuring business cycles*, burn46-1, National bureau of economic research.

CHAMBERLAIN, G. AND M. ROTHSCHILD (1982): "Arbitrage, factor structure, and mean-variance analysis on large asset markets," National Bureau of Economic Research Cambridge, Mass., USA.

——— (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304.

CHANG, W.-J., S. J. MONAHAN, A. OUAZAD, AND F. VASVARI (2021): "The higher moments of future earnings," *Accounting Review*, 96, 91–116.

FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): "The Generalized Dynamic-Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82, 540–554.

FORNI, M. AND M. LIPPI (2001): "The Generalized Dynamic-Factor Model: Representation Theory," *Econometric Theory*, 17, 1113–1141.

GALLANT, R. A. AND G. TAUCHEN (1996): "Which Moments to Match?" *Econometric Theory*, 1996, 657–681.

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.

GUVENAN, F., S. OZKAN, AND J. SONG (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122, 621–660.

GUVENAN, F., S. SCHULHOFER-WOHL, J. SONG, AND M. YOGO (2017): "Worker Betas: Five Facts about Systematic Earnings Risk," *American Economic Review: Papers Proceedings*, 107, 398–403.

HANSEN, B. (2022): *Econometrics*, Princeton University Press.

HANSEN, L. P. (2012): "Dynamic valuation decomposition within stochastic economies," *Econometrica*, 80, 911–967.

HEATHCOTE, J., F. PERRI, AND G. L. VIOLANTE (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic Dynamics*, 13, 15–51.

HEATHCOTE, J., F. PERRI, G. L. VIOLANTE, AND L. ZHANG (2023): "More Unequal We Stand? Inequality Dynamics in the United States, 1967–2021," *Review of Economic Dynamics*, 50, 235–266.

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.

IAO, M. C. AND Y. J. SELVAKUMAR (2024): "Estimating HANK with Micro Data," Tech. rep., New York University.

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.

LJUNGQVIST, L. AND T. J. SARGENT (2018): *Recursive Macroeconomic Theory*, MIT Press, 4 ed.

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

MEYER, B. D. AND J. X. SULLIVAN (2013): "Consumption and Income Inequality and the Great Recession," *American Economic Review: Papers Proceedings*, 103, 178–183.

——— (2023): "Consumption and Income Inequality in the United States since the 1960s," *Journal of Political Economy*, 131, 247–284.

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

PARKER, J. A. AND A. VISSING-JORGENSEN (2009): "Who Bears Aggregate Fluctuations and How?" *American Economic Review: Papers Proceedings*, 99, 399–405.

PIKETTY, T. AND E. SAEZ (2003): "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, 118, 1–39.

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.

STOCK, J. AND M. WATSON (1988): "A Probability Model of the Coincident Economic Indicators," *NBER Working Paper Series*.

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.

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.