Linearizing stochastic systems with the Koopman Operator*

Yatheesan J. Selvakumar

New York University (NYU)

January 22, 2024

Abstract

This paper provides an alternative computationally-efficient approach to approximate non-linear stochastic models. The novel method combines both perturbation and projection methods and draws on results from Koopman operator theory. I state convergence results and demonstrate its capabilities on a highly non-linear consumption-based asset-pricing model with external habits.

Keywords: Koopman operator, linear approximation, first-order approximations, hybrid projection-perturbation methods, asset pricing, external habits

JEL Classification Numbers: ...

*I thank Thomas Sargent, Timothy Christensen, Simon Gilchrist, Jarda Borovicka, Mikkel Plagborg-Moller, Juan Rubio-Ramirez, Ivan Petrella, Balint Szoke, Francesco Furno, Brandon Kaplowitz, Luca Gagliardone and Christopher Adjaho for helpful discussions and comments.

1 Introduction

Economists have long used linear approximations to ease the complexity of solving, estimating and controlling non-linear models of the economy. First-order perturbation methods have become popular since they admit the use of simple, robust tools for analyses[1], even though their "local" nature of can forego accuracy in certain instances. The perceived importance of this drawback is revealed in the advent of methods that include higher-order terms[2]. While these models improved on the accuracy, they inherently lose the convenience associated with the first-order approximation[3]. This paper aims to fill the gap by providing a methodology, called the **Koopman Linearization**, for first-order *global* approximations of non-linear stochastic models. The "first-order" aspect retains the convenience of linear approximations, while the "global" aspect means that it better approximates non-linearities.

It is useful to think about the Koopman Linearization as a perturbation followed by a projection. The perturbation is a first-order small-noise expansion of the model that splits the non-linear stochastic dynamics into a deterministic part that is non-linear in the functions of the states; and a stochastic part that is linear in the shocks. The second part is based on the Koopman operator[4] that *linearly* represents the non-linear deterministic dynamics of functions of the states. Since it is infinite dimensional, a projection is used to approximate the Koopman operator as a linear combination of functions of an orthonormal basis. In this sense, the Koopman Linearization brings together perturbation and projection in a unique way[5].

[1] See Fernandez-Villaverde et al. (2007) and Aruoba et al. (2017) for such examples

[2] For example, Schmitt-Grohe and Uribe (2002) introduce second-order terms in the linearizations while Kim et al. (2008), Andreasen et al. (2018) and others proposed refined versions of such approximations. Kim and Kim (2003) also present an example that shows the clear need for higher-order approximations

[3] In response, Aruoba et al. (2017) suggest new vector auto-regression models that included higher order terms.

[4] See Koopmans (1947)

[5] Judd (1996) discusses the promise of combining projection and perturbation methods. Its value essentially boils down to the observation that both methods are complementary in their strengths and weaknesses. See Section 11 in Judd (1996) for details

After describing the methodology, I provide convergence results of the Koopman approximation and quantify the bias associated with the small-noise approximation. Then, I demonstrate the Koopman Linearization by applying it to the well-known consumption habits model of Campbell and Cochrane (1999) in Section 3. Due to the kink of the parametrized surplus consumption function, the method prevails even when standard first-order methods are not applicable. I show evidence of high accuracy in terms of impulse responses and via simulation. Finally, I combine the Koopman Linearization with the Kalman filter to estimate hidden states and structural parameters via maximum likelihood.

2 The Koopman Linearization

Let $\mathbf{x}_t$ be a $n \times 1$ vector of state variables and $\mathbf{y}_t$ be a $m \times 1$ vector of observables. The joint dynamics of $(\mathbf{x}_t, \mathbf{y}_t)$ are given by the non-linear state-space model

$$\mathbf{x}_{t+1} = \mathbf{T}(\mathbf{x}_t, \boldsymbol{\sigma}\boldsymbol{\varepsilon}_{t+1}) \tag{1}$$

$$\mathbf{y}_t = \mathbf{F}(\mathbf{x}_t) + \mathbf{u}_t \tag{2}$$

where $\boldsymbol{\varepsilon}_{t+1} \in \mathbb{R}^l$ and $\mathbf{u}_t \in \mathbb{R}^m$ vectors of standard Gaussian shocks. To ensure existence and uniqueness of a fixed point $\mathbf{x}^*$, $\mathbf{T} : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}^k$ is invertible and Lipschitz continuous[6]. Finally, let $\mathbf{F} = [f_1, f_2, \ldots, f_m]^\top$ be a vector of L^2 functionals, where $f : \mathbb{R}^n \to \mathbb{R}$.[7] Within the context of economic models, we assume that (1)-(2) is the *equilibrium* dynamics of the model.

Fix a vector of L^2 functions $\mathbf{g} = [g_1, g_2, \ldots, g_k]^\top$, where $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, k$.

[6] See Mezic (2022)

[7] f is an L^2 function if it is square-integrable, i.e. $\int_X |f|^2 d\mu$ with respect to measure μ exists

Applying $\mathbf{g}$ to both sides of (1) and taking a first order expansion around $\boldsymbol{\sigma} = \mathbf{0}$ yields

$$\mathbf{g}(\mathbf{x}_{t+1}) = \mathbf{g}(\mathbf{T}(\mathbf{x}_t, \mathbf{0})) + \mathbf{B}_{\mathbf{g}}(\mathbf{x}_t)\boldsymbol{\sigma}\boldsymbol{\varepsilon}_{t+1} + \mathcal{O}(||\boldsymbol{\sigma}||^2) \tag{3}$$

where $\mathbf{B}_{\mathbf{g}}(\mathbf{x}_t) = \mathcal{J}_{\mathbf{g}}(\mathbf{T}(\mathbf{x}_t, \mathbf{0}))\, \mathbf{T}_2(\mathbf{x}_t, \mathbf{0})$. The matrix $\mathcal{J}_{\mathbf{g}} \in \mathbb{R}^{k \times n}$ denotes the Jacobian of $\mathbf{g}$, and $\mathbf{T}_2 \in \mathbb{R}^{n \times l}$ denotes the derivative of $\mathbf{T}$ with respect to the second argument. The first term in (3) is the deterministic (or zero-th order) part of (1), and the second term approximates the impact of shocks. Thus, the expansion can be seen as a perturbation around the deterministic model, as is commonly described in the economic literature[8] That the expansion is not taken around $\mathbf{x}^*$ is an attractive feature, since the approximation accuracy is of order $||\,\boldsymbol{\sigma}\,||^2$, not $||\,\mathbf{x}_t - \mathbf{x}^*\,||^2$ as with other perturbation methods. The other attractive feature is that the perturbation does not require differentiability in the first argument of $\mathbf{T}$. The approximate law of motion (3) is linear in $\boldsymbol{\varepsilon}_{t+1}$ and non-linear in $\mathbf{x}_t$. In the next section we use the Koopman operator to compute a linear approximation of $\mathbf{g}(\mathbf{T}(\mathbf{x}_t, \mathbf{0}))$.

2.1 The Koopman Operator

Let $\mathcal{K} : L^2 \to L^2$ be the composition operator that maps $\mathcal{K} : \mathbf{g} \mapsto \mathbf{g} \circ \mathbf{T}(\cdot, \mathbf{0})$. Therefore

$$\mathbf{g}(\mathbf{T}(\mathbf{x}_t, \mathbf{0})) = \mathcal{K}\,\mathbf{g}(\mathbf{x}_t)$$

Then equation (3) can be rewritten accordingly as

$$\mathbf{g}(\mathbf{x}_{t+1}) = \mathcal{K}\,\mathbf{g}(\mathbf{x}_t) + \mathbf{B}_{\mathbf{g}}(\mathbf{x}_t)\,\boldsymbol{\sigma}\,\boldsymbol{\varepsilon}_{t+1} + \mathcal{O}(||\,\boldsymbol{\sigma}\,||^2) \tag{4}$$

By virtue of it being the composition operator, $\mathcal{K}$ is linear, making the law of

[8] See Bhandari et al. (2021) or Borovicka and Hansen (2014) for applications of the small-noise expansion in economic settings. The important difference is that this method approximates the fully non-linear equilibrium solution while other papers in the literature take a small-noise expansion and then solve the approximate model.

motion of quasi-linear in $\mathbf{g}(\mathbf{x}_t)$. Since $\mathcal{K}$ acts on a function space, is therefore infinite dimensional, which poses a practical problem for implementation. In section 2.4 below, I propose an accessible approach for approximating $\mathcal{K}$ with finite dimensional matrices. Before that, I fix ideas with an example of the Koopman operator in a simple deterministic setting.

2.2 Illustrative example

Define $\mathbf{x}_t = [x_{1,t}, x_{2,t}]^\top$, and assume it follows deterministic dynamics

$$\begin{aligned} x_{1,t+1} &= \mu x_{1,t} \\ x_{2,t+1} &= \lambda x_{2,t} + \rho x_{1,t}^2 \end{aligned} \tag{5}$$

Choose judiciously $\mathbf{g}(\mathbf{x}_t) = [g_1(\mathbf{x}_t), g_2(\mathbf{x}_t), g_3(\mathbf{x}_t)]^\top = [x_{1,t}, x_{2,t}, x_{1,t}^2]^\top$.

The dynamics of (5) can be rewritten in matrix form as

$$\underbrace{\begin{bmatrix} g_1(\mathbf{x}_{t+1}) \\ g_2(\mathbf{x}_{t+1}) \\ g_3(\mathbf{x}_{t+1}) \end{bmatrix}}_{\mathbf{g}(\mathbf{x}_{t+1})} = \underbrace{\begin{bmatrix} \mu & 0 & 0 \\ 0 & \lambda & \rho \\ 0 & 0 & \mu^2 \end{bmatrix}}_{\mathcal{K}} \underbrace{\begin{bmatrix} g_1(\mathbf{x}_t) \\ g_2(\mathbf{x}_t) \\ g_3(\mathbf{x}_t) \end{bmatrix}}_{\mathbf{g}(\mathbf{x}_t)}$$

In this simple example, $\mathcal{K}$ specializes to a 3×3 matrix. Notice that there is no approximation error – by transforming the dynamics of $\mathbf{x}_t$ into the space of $\mathbf{g}(\mathbf{x}_t)$, the non-linear model of (5) has a *linear* representation. This is true only for the particular chosen $\mathbf{g}$, which is easy to choose for this simple problem. In the next section, I propose a systematic choice of $\mathbf{g}$ that is useful in more general situations.

2.3 Finite Dimensional $\mathcal{K}$

The ability to approximate the Koopman operator with a finite dimensional matrix is a key step in operationalizing the Koopman Linearization. An important concept to this end is Koopman invariant subspaces.

Definition 1. *Let* $\mathcal{G} := \text{span}\{g_1, \ldots, g_N\}$ *be the subspace spanned by functionals* $g_1, \ldots, g_N \in L^2(\mathcal{X})$. $\mathcal{G}$ *is* ***Koopman invariant*** *if for any* $\gamma \in \mathcal{G}$, $\mathcal{K}\,\gamma \in \mathcal{G}$.

Any set of basis functions that form a Koopman invariant subspace give rise to a finite dimensional $\mathcal{K}$. Reverting back to the example 2.2 reveals one instance of this. Define the subspace $\mathcal{G}_3 := \text{span}\{g_1, g_2, g_3\}$. Then

$$\begin{aligned}
\mathcal{K}\, g_1(\mathbf{x}_t) &= g_1(\mathbf{x}_{t+1}) = \mu x_{1,t} = \mu g_1(\mathbf{x}_t) \in \mathcal{G}_3 \\
\mathcal{K}\, g_2(\mathbf{x}_t) &= g_2(\mathbf{x}_{t+1}) = \lambda x_{2,t} + \rho x_{1,t}^2 = \lambda g_2(\mathbf{x}_t) + \rho g_3(\mathbf{x}_t) \in \mathcal{G}_3 \\
\mathcal{K}\, g_3(\mathbf{x}_t) &= g_3(\mathbf{x}_{t+1}) = \mu^2 x_{1,t}^2 = \mu^2 g_3(\mathbf{x}_t) \in \mathcal{G}_3
\end{aligned}$$

and so $\mathcal{G}_3$ is clearly a Koopman invariant subspace. Indeed, $\mathcal{K}$ is a (finite-dimensional) matrix.

Koopman eigenfunctions. Let $\{\varphi_i\}_{i=1}^{\infty}$ be the set of (infinitely many) eigenfunctions of $\mathcal{K}$, with associated eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$. Brunton et al. (2015) shows that any finite subset of eigenfunctions form a Koopman invariant subspace. Moreover, the zero-th order (deterministic) dynamics of $\varphi_i(\mathbf{x}_t)$ is linear with coefficient λ_i.

$$\varphi_i(\mathbf{T}(\mathbf{x}_t, \mathbf{0})) = \mathcal{K}\,\varphi_i(\mathbf{x}_t) = \lambda_i \varphi_i(\mathbf{x}_t) \quad \text{for all } i = 1, \ldots \tag{6}$$

Suppose we choose N eigenfunctions and stack them to form $\boldsymbol{\varphi} = [\varphi_1, \ldots, \varphi_N]^\top$.

The dynamics of $\boldsymbol{\varphi}(\mathbf{x}_t)$ is

$$\boldsymbol{\varphi}(\mathbf{T}(\mathbf{x}_t, \mathbf{0})) = \boldsymbol{\Lambda}\, \boldsymbol{\varphi}(\mathbf{x}_t) \tag{7}$$

where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_N)$. Clearly the finite-dimensional Koopman operator associated with $\boldsymbol{\varphi}$ is $\boldsymbol{\Lambda}$.

2.4 Approximate Koopman operators

The previous subsection shows that $\mathcal{K}$ associated with any subset of its eigenfunctions is finite dimensional. The question then becomes: how do we find these eigenfunctions? One common method is to approximate $\mathcal{K}$ via projection. Williams et al. (2015) refers to this as the Extended Dynamic Mode Decomposition method[9]

Let the $n \times 1$ vector $\mathbf{x}_i$ be a point in the discretized state-space $\mathcal{X}$, and create the $n \times M$ matrices

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M], \quad \mathbf{X}' = [\mathbf{T}(\mathbf{x}_1, \mathbf{0}), \ldots, \mathbf{T}(\mathbf{x}_M, \mathbf{0})] \tag{8}$$

Let $\{g_i\}_{i=1}^{\infty}$ form an orthonormal basis of L^2, for example the set of Chebyshev polynomials[10]. Choose a subset of $N < M$ functionals and stack them to form the $M \times 1$ vector $\mathbf{g} = [g_1, \ldots, g_N]^\top$. Then, create the $N \times M$ data matrices

$$\mathbf{g}(\mathbf{X}) = \begin{bmatrix} g_1(\mathbf{x}_1) & \ldots & g_1(\mathbf{x}_M) \\ \vdots & \ldots & \vdots \\ g_N(\mathbf{x}_1) & \ldots & g_N(\mathbf{x}_M) \end{bmatrix}, \quad \mathbf{g}(\mathbf{X}') = \begin{bmatrix} g_1(\mathbf{T}(\mathbf{x}_1, \mathbf{0})) & \ldots & g_1(\mathbf{T}(\mathbf{x}_M), \mathbf{0}) \\ \vdots & \ldots & \vdots \\ g_N(\mathbf{T}(\mathbf{x}_1, \mathbf{0})) & \ldots & g_N(\mathbf{T}(\mathbf{x}_M, \mathbf{0})) \end{bmatrix}$$

[9] Under the assumptions below, the Williams et al. (2015) show below that the EDMD method to the numerical approximation obtained from a Galerkin method.

[10] Christensen (2017) uses a sieve estimator to approximate the Perron-Frobenius eigenfunctions from data in a similar vein, and estabilishes consistency and convergence rates. The Perron-Frobenius operator is the left-adjoint of the Koopman operator.

Approximate Koopman operator $\widehat{\mathbf{K}}$, is the solution to

$$\widehat{\mathbf{K}} = \arg \min_{K \in \mathbb{R}^{N \times N}} \|\mathbf{g}(\mathbf{X}') - K\,\mathbf{g}(\mathbf{X})\|_F^2 \implies \widehat{\mathbf{K}} = \mathbf{g}(\mathbf{X}')\,\mathbf{g}(\mathbf{X})^\dagger$$

where $\|\cdot\|_F$ is the Frobenius norm and $(\cdot)^\dagger$ is the generalized inverse.[11]

Then approximate Koopman eigenfunctions are given by $\widehat{\varphi}_i = \widehat{\phi}_i\,\mathbf{g}$ for $i = 1, \ldots, N$ where $\widehat{\phi}_i$ is the ith left eigenvector of $\widehat{\mathbf{K}}$ with associated eigenvalue $\widehat{\lambda}_i$. Brunton et al. (2015) also show that $\widehat{\varphi}_i(\mathbf{x}_t)$ has linear (deterministic) dynamics with coefficient λ_i, since

$$\widehat{\varphi}_i(\mathbf{T}(\mathbf{x}_t, \mathbf{0})) = \widehat{\phi}_i\,\mathbf{g}(\mathbf{T}(\mathbf{x}_t, \mathbf{0})) = \widehat{\phi}_i\,\mathcal{K}\,\mathbf{g}(\mathbf{x}_t) = \lambda_i \widehat{\phi}\,\mathbf{g}(\mathbf{x}_t) = \lambda_i \widehat{\varphi}_i(\mathbf{x}_t) \tag{9}$$

where the second equality is the definition of $\mathcal{K}$ and the third equality uses the definition of a left eigenvector.

Just as with (7), stacking the eigenfunctions $\widehat{\boldsymbol{\varphi}}(\mathbf{x}_t) := [\widehat{\phi}_1, \ldots, \widehat{\phi}_N]\,\mathbf{g}(\mathbf{x}_t)$ and $\widehat{\boldsymbol{\Lambda}} := \text{diag}(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_N)$ yields the linear (deterministic) dynamics

$$\widehat{\boldsymbol{\varphi}}(\mathbf{T}(\mathbf{x}_t, \mathbf{0})) = \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\varphi}}(\mathbf{x}_t)$$

2.5 Measurement equation

In finding a linear approximation to the dynamics of state transition equation (1), the state variables have been transformed from $\mathbf{x}_t$ to $\widehat{\boldsymbol{\varphi}}(\mathbf{x}_t)$. To then find a linear approximation to the measurement equation, compute a matrix $\mathbf{H}$ that solves

$$\arg \min_{H \in \mathbb{R}^{m \times N}} ||\,\mathbf{F}(\mathbf{X}) - H\widehat{\boldsymbol{\varphi}}(\mathbf{X})||_F^2 \tag{10}$$

[11] Since $N < M$, $\widehat{\mathbf{K}}$ is well-defined and equivalent to the standard formula for least-squares regression.

The approximation measurement equation is therefore

$$\mathbf{y}_t = \mathbf{H}\,\widehat{\boldsymbol{\varphi}}(\mathbf{x}_t) + \mathbf{u}_t \tag{11}$$

and the system In the next section, I provide pseudo-code to implement the Koopman Linearization of non-linear state-space model (1)-(2).

2.6 Pseudo-code for Koopman Linearization

1. Choose order of Chebyshev polynomial, N

$$\mathbf{g}(s) = [g_0(s), \ldots, g_N(s)]^\top \tag{12}$$

   where $g_i(\mathbf{s})$ is chebyshev polynomial of order i

2. Create matrices $\mathbf{X}$ and $\mathbf{x}'$ of discretized state-space $\mathbf{X}$ as in (8)

3. Calculate $\widehat{\mathbf{K}}$ via projection

$$\widehat{\mathbf{K}} = \mathbf{g}(\mathbf{X}')\,\mathbf{g}(\mathbf{X})^\dagger \tag{13}$$

4. Compute N eigenvector-value pairs of $\widehat{\mathbf{K}}$, $\{\widehat{\phi}_i, \widehat{\lambda}_i\}_{i=0}^N$ and approximate eigenfunctions and eigenvalues

$$\widehat{\boldsymbol{\varphi}}(\mathbf{s}) = [\widehat{\phi}_0, \ldots, \widehat{\phi}_N]\,\mathbf{g}(s)$$
$$\boldsymbol{\Lambda} = \text{diag}(\widehat{\lambda}_0, \ldots, \widehat{\lambda}_N)$$

5. Calculate conditional covariance function, $\widehat{\mathbf{B}}_{\boldsymbol{\varphi}}(\mathbf{x}) = \nabla\widehat{\boldsymbol{\varphi}}(\mathbf{x})\,\mathbf{T}_2(\mathbf{x}, 0)$

6. Calculate measurement approximation via projection

$$\mathbf{H} = \mathbf{F}(\mathbf{X})\widehat{\boldsymbol{\varphi}}(\mathbf{X})^{\dagger} \tag{14}$$

7. The Koopman Linearized (KL) model is

$$\widehat{\boldsymbol{\varphi}}(\mathbf{x}_{t+1}) \approx \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\varphi}}(\mathbf{x}_t) + \widehat{\mathbf{B}}_{\boldsymbol{\varphi}}(\mathbf{x}_t)\,\boldsymbol{\varepsilon}_{t+1} \tag{15}$$

$$\mathbf{y}_t \approx \mathbf{H}\,\widehat{\boldsymbol{\varphi}}(\mathbf{x}_t) + \mathbf{u}_t \tag{16}$$

2.7 Approximation error

There are three sources of approximation error for the Koopman Linearization: The Koopman operator approximation, the measurement equation approximation, the perturbation approximation.

For the **Koopman operator approximation error**, I borrow results from Korda and Mezic (2017) who show that under regularity conditions that our framework satisfies $\widehat{\mathbf{K}}$ converges to $\mathcal{K}$ in the strong operator topology as the order of the chebyshev polynomials becomes large ($N \to \infty$) and the number of discretization points becomes large ($M \to \infty$).[12,13]. The authors also show under the same assumptions that the eigenpairs of $\widehat{\mathbf{K}}$ ($\hat{\lambda}_i, \hat{\varphi}_i$) weakly converge to the eigenpairs of $\mathcal{K}$ (λ_i, φ_i for $i = 1, \ldots, N$)[14].

The **measurement equation approximation error** is $\mathbf{F}(\mathbf{x}_t) - \mathbf{H}\,\boldsymbol{\varphi}(\mathbf{x}_t)$. Since $=$

[12]A sequence of bounded operators $\mathbf{K}_N : L^2 \to L^2$ defined on a Hilbert space L^2 convergences strongly (or in the strong operator topology) to an operator $\mathcal{K} : L^2 \to L^2$ if

$$\lim_{M,N\to\infty} ||\,\mathbf{K}_N\, g\, \mathcal{K}\, g|| \quad \forall g \in L^2 \tag{17}$$

[13]The assumptions required for convergence are that 1) $\mathcal{K}$ is bounded and 2) $g_1, \ldots, g_N$ are selected from an orthonormal basis of L^2. See Korda and Mezic (2017) for further details

[14]A sequence of elements $f_N \in L^2$ of a Hilbert space L^2 converges weakly to $f \in L^2$ if

$$\lim_{N\to\infty} \langle f_N, g\rangle = \langle f, g\rangle \tag{18}$$

$[f_1, \ldots, f_m]^\top$ is a vector of L^2 functionals and $\boldsymbol{\varphi}$ is the linear combination of chebyshev polynomials $g_1, \ldots, g_N$ are a subset of an orthonormal basis for L^2, as the order of the chebyshev polynomials (N) and the number of discretization points become large (M),

$$\lim_{M,N\to\infty} |f_i(\mathbf{x}_t) - \mathbf{H}_N^{(i,\cdot)} \boldsymbol{\varphi}_N(\mathbf{x}_t)| = 0 \quad i = 1, \ldots, m \tag{19}$$

for the i-th row $\mathbf{H}^{(i,\cdot)}$ of the matrix $\mathbf{H}_N$, and where we make explicit the dependence on N for clarity.

The **perturbation approximation error** is shown in (4). This is the only error does not vanish for an increasing N or M and is the root of any asymptotic bias in the approximation. The error is of order $||\boldsymbol{\sigma}||^2$, and only vanishes if $\boldsymbol{\sigma} = \mathbf{0}$. For small values of $\boldsymbol{\sigma}$, the approximation error is therefore small, which is the usual assumption in the small-noise expansion literature.

3 Asset pricing with external habits

We apply the Koopman Linearization methodology to the model of external habits of Campbell and Cochrane (1999). A representative agent maximises

$$\mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \frac{(C_t - X_t)^{1-\gamma}}{1-\gamma} \right]$$

with coefficient of risk aversion γ, level of habit X_t and subjective discount factor β. A convenient object is the surplus consumption ratio $S_t = \frac{C_t - X_t}{C_t}$. Following the literature, I assume log consumption growth is an i.i.d process with mean μ_c, and parameterize the log surplus consumption ratio as

$$s_{t+1} = (1-\phi)\bar{s} + \phi s_t + \lambda(s_t)\epsilon_{t+1} \tag{20}$$

where $\epsilon_{t+1} = \Delta c_{t+1} - \mu_c$, and define $\lambda(s_t)$ below.

The intertemporal marginal rate of substitution for the agent is

$$m_{t+1} = \log(\beta) - \gamma[\mu_c - (1-\phi)(s_t - \bar{s}) + (1+\lambda(s_t))\epsilon_{t+1}]$$

I assume that the risk-free rate is observed with noise, so the measurement equation is

$$\begin{aligned} r_t^f &= r_t^{f*} + \sigma_u u_t \\ &= -\mathbb{E}_t[m_{t+1}] + \frac{1}{2}\sigma_{m,t}^2 + \sigma_u u_t \\ &= -\log\beta + \gamma\mu_c - \gamma(1-\phi)(s_t - \bar{s}) - \frac{1}{2}\gamma^2[1+\lambda(s_t)]^2\sigma_c^2 + \sigma_u u_t \end{aligned} \tag{21}$$

To engineer key features in the model consistent with the empirical literature, I follow Wachter (2005) and define $\lambda(s_t)$

$$\lambda(s_t) = \begin{cases} \frac{1}{\bar{S}}\sqrt{1-2(s_t-\bar{s})} - 1, & \text{if } s \leq \bar{s} \\ 0, & \text{otherwise} \end{cases}$$
$$\bar{S} = e^{\bar{s}} = \sigma_c\sqrt{\frac{\gamma}{1-\phi-b/\gamma}}$$

The non-linear aspect of λ is significant. First, $\lambda(s_t)$ is kinked at $\bar{s}$ making a simple log-linear approximation unsuitable. Furthermore $\lambda(s_t)$ is non-linear even when $s < \bar{s}$. Put together equations (20)-(21) define the non-linear state-space model.

I follow Wachter (2005) in parameterizing the model, setting $\mu_c = \frac{2.2}{400}, \sigma_c = \frac{0.86}{200}, \gamma = 2.0, b = 0.011, \phi = 0.89^{\frac{1}{4}}, \beta = 0.93^{\frac{1}{4}}$. Finally, I set standard deviation of the measurement error to $\sigma_u = 0.001$.

To demonstrate the Koopman Linearization, I implement the pseudo-code in section 2.6, setting $N = 5$. Figure 1 plots impulse responses of s_t and r_t^f to a one standard deviation increase in ε_{t+1} for both the true model and the approximate model. Since

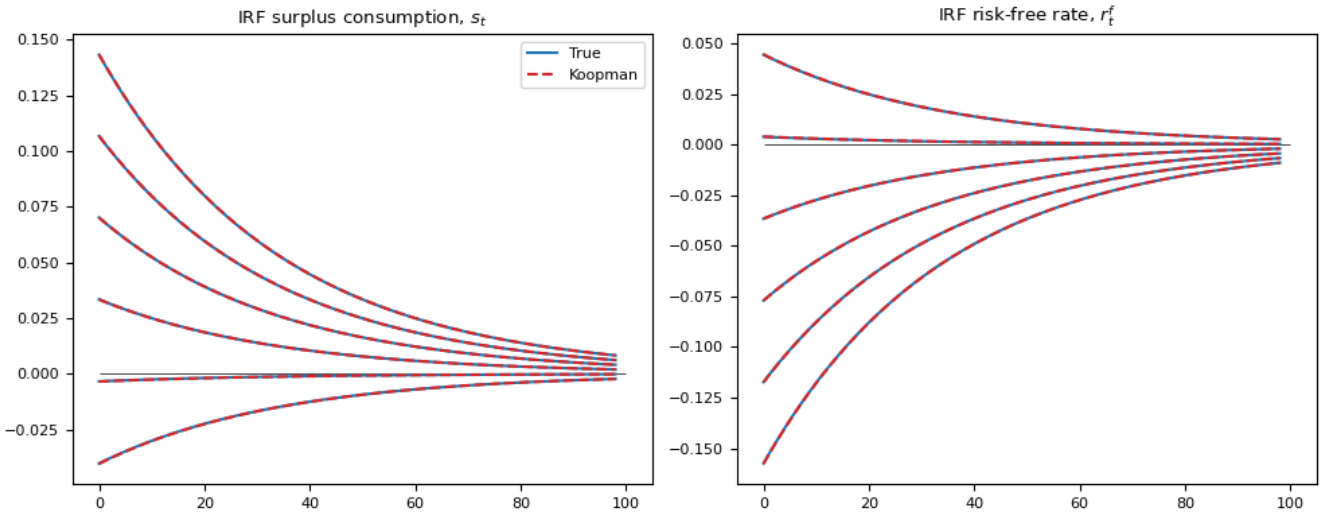

Figure 1: Non-linear impulse response approximation

the model is non-linear, I calculate the impulse response at different initial values between $\pm 10\%$ of $\bar{s}$. The figure suggests high accuracy in approximation in response to a moderately sized-shock at many different initial values.

I further test the accuracy of the approximation using the accuracy metric of Den Haan (2010). I simulate J paths of shocks $\{\varepsilon_{t+1}^j\}_{t=0}^T$ and feed the shocks into both the true model and the KL model. For path j and time t, define the approximation error of the log surplus consumption ratio

$$\mathcal{E}_t^j(s) = s_t^j - \hat{s}_t^j$$

where $\hat{s}_t^j$ denotes the approximate state from the KL model. The approximation error for the r_t^f is defined analogously. Table 1 reports the mean, standard deviation and maximum-absolute error for $J = 500$ and $T = 1000$, suggesting a high degree of accuracy for the KL approximation.

Object	Mean	Standard deviation	Max-absolute error
$\mathcal{E}_t^j(s)$	$-6.1e^{-5}$	$5e^{-4}$	$1e^{-1}$
$\mathcal{E}_t^j(f)$	$-9.9e^{-7}$	$7.5e^{-5}$	$2.6e^{-4}$

Table 1: Accuracy metrics for simulated data

3.1 Estimation

A powerful feature of the KL is that it opens the door to using existing linear techniques on highly non-linear systems. Two immediate applications are hidden-state estimation and likelihood-based parameter estimation via the Kalman filter.

One way to implement this is by replacing the Koopman Linearized system in (15) - (16) with

$$\widehat{\boldsymbol{\varphi}}(\mathbf{x}_{t+1}) \approx \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\varphi}}(\mathbf{x}_t) + \widehat{\mathbf{B}}(\mathbf{x}^*)\,\boldsymbol{\varepsilon}_{t+1} \tag{22}$$

$$\mathbf{y}_t \approx \mathbf{H}\,\widehat{\boldsymbol{\varphi}}(\mathbf{x}_t) + \mathbf{u}_t \tag{23}$$

where we have replaced $\widehat{\mathbf{B}}(\mathbf{x}_t)$ with its steady-state analogue $\widehat{\mathbf{B}}(\mathbf{x}^*)$. This approximation is crucial in mapping the approximate model into the state-space form required with the Kalman filter.[15] The downside to this approximation is that it introduces an additional approximation error of size $(\widehat{\mathbf{B}}(\mathbf{x}_t) - \widehat{\mathbf{B}}(\mathbf{x}^*))\,\boldsymbol{\varepsilon}_{t+1}$.

We simulate a time-series ($T = 1000$) of log surplus consumption ratio and risk-free rate. We assume that only the risk-free rate is observed and filter the hidden states from the Koopman Linearized state-space system (22) using the Kalman filter. Figure 2 plots the time-series of the simulated states in blue, the KL mean estimate of the states using the Kalman filter in red, and the mean estimate of the filtered states using the adaptive particle filter of Herbst and Schorfheide (2016) with 1000 particles. The figure

[15] Without this approximation, equation (15) has a stochastic covariance term that depends on $\widehat{\varphi}(\mathbf{x}_t)$. Existing filters, such as the ensemble filter (see Evensen (1994) and Houtekamer (1998)) are flexible enough to capture these kinds of properties of a state-space model.

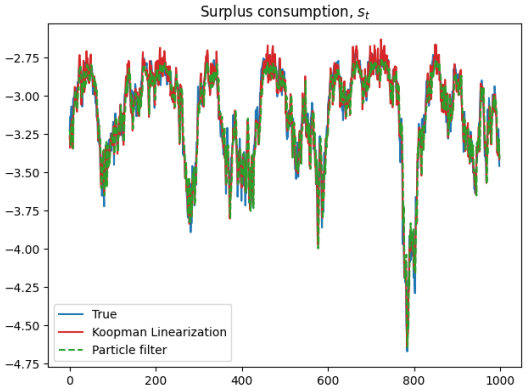

Figure 2: Simulated path of states

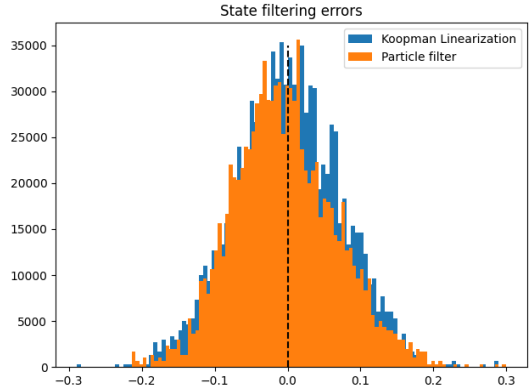

Figure 3: Distribution of filtering errors

shows that the KL states that those estimated by the particle filter are very close. In some instances, the Koopman states are more volatile than either the true of PF states.

Studying the differences systematically, I compute J of these paths. Then for each path j, I calculate errors of the estimated-states for both the KL and the particle filter. Recycling notation, define

$$\mathcal{E}^{KL}_{j,t} = s_{j,t} - \hat{s}^{KL}_{j,t}$$
$$\mathcal{E}^{PF}_{j,t} = s^{j}_{t} - \hat{s}^{PF}_{j,t}$$

Figure 3 plots the distribution of $\mathcal{E}^{KL}_{j,t}$ and $\mathcal{E}^{PF}_{j,t}$ from our simulated paths. The error distributions appear to be similar. The KL errors have mean 0.001 with standard deviation 0.073 and the PF errors have mean -0.003 and standard deviation 0.070. The KL is much more computationally efficient, however: the filtering for $J = 1000$ and $T = 1000$ took the particle filter 23 minutes, and it took the KL filter 13 minutes, a 76% improvement[16].

Next we study the distribution of the maximum likelihood estimate associated with the KL. The literature has historically resorted to either calibration or method

[16]The monte carlo exercise was executed on an Apple M1 Pro with 16GB memory, 8 performance cores and 2 efficient cores

of moments techniques to estimate Campbell and Cochrane (1999). For example, Tallarini and Zhang (2005) who use the efficient methods of moments strategy of Gallant and Tauchen (1996), while Chang et al. (2005) and Engsted and Moller (2010) follow generalized method of moments procedures to estimate parameters. I instead estimate Koopman Linearized model via maximum likelihood[17].

As before, we generate $J = 1000$ paths of length $T = 1000$, and calculate the maximum likelihood estimates for each parameter, $\mu_c, \sigma_c, \gamma, b, \phi, \beta$. Figure 4 plots the histograms associated with finite-sample distribution for each MLE parameter, and the red dotted line denotes the true parameter value generating the simulated data. Table 2 shows the mean of the estimated parameters across the monte carlo samples. For the majority of the parameters, the mean estimate of the parameters is close to the true value. The only real outlier is σ_c, where the mean estimate is significantly larger from the true parameter.

	μ_c	σ_c	γ	b	ϕ	β
True value	0.0055	0.0043	2.0	0.011	0.971	0.982
Sample mean of MC draws	0.0055	0.0077	1.97	0.011	0.971	0.982

Table 2: Mean parameter estimate from 1000 MC paths

[17]This approach can be viewed similar in spirit to Gallant and Tauchen (1996) by consider the the KL model as an auxiliary model

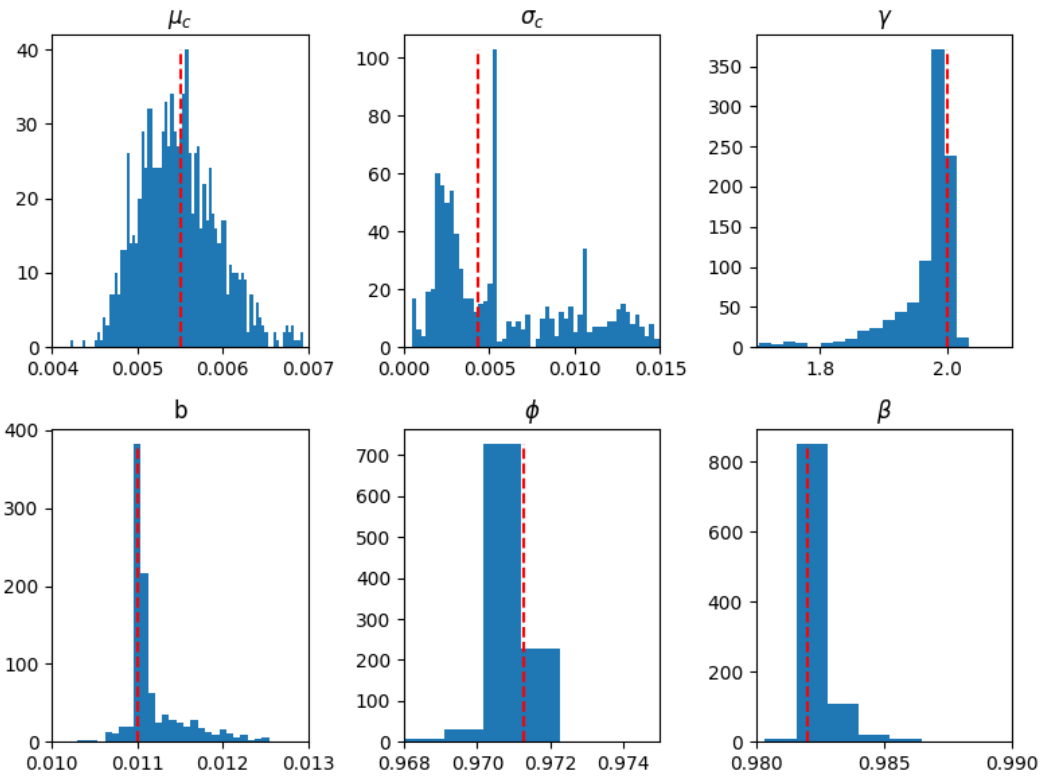

Figure 4: MLE sampling distribution ($J = 1000$)

4 Concluding remarks

This paper presents a methodology for globally approximating non-linear equilibrium models that retains the convenience associated with a linear framework. The algorithm combines projection and perturbation methods and extends Koopman Operator theory to the stochastic case. I characterize the approximation quality and provide results on convergence, drawing on existing literature. Then, I demonstrate the methods capabilities on a model of external habits for which standard linear approximations are unsuitable, by applying it to the problem of filtering and structural parameter estimation.

References

ANDREASEN, M. M., J. FERNÁNDEZ-VILLAVERDE, AND J. F. RUBIO-RAMÍREZ (2018): "The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications," *The Review of Economic Studies*, 85, 1–49.

ARUOBA, S. B., L. BOCOLA, AND F. SCHORFHEIDE (2017): "Assessing DSGE model nonlinearities," *Journal of Economic Dynamics and Control*, 83, 34–54.

BHANDARI, A., D. EVANS, M. GOLOSOV, AND T. J. SARGENT (2021): "Inequality, Business Cycles, and Monetary-Fiscal Policy," *Econometrica*, 89, 2559–2599.

BOROVICKA, J. AND L. P. HANSEN (2014): "Robust Preference Expansions," *Unpublished Working Paper*.

BRUNTON, S., B. BRUNTON, J. PROCTOR, AND N. KUTZ (2015): "Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control," *arXiv*.

CAMPBELL, J. Y. AND J. H. COCHRANE (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behaviour," *Journal of Political Economy*, 107, 205–51.

CHANG, Y., S. KIM, AND J. Y. PARK (2005): "GMM estimation of an asset pricing model with havit persistence," Tech. rep., university of Chicago Economioc s.

CHRISTENSEN, T. M. (2017): "Nonparametric Stochastic Discount Factor Decomposition," *Econometrica*, 85, 1501–1536.

DEN HAAN, W. (2010): "Comparison of solutions to the incomplete markets model with aggregate uncertainty," *Journal of Economic Dynamics Control*, 34, 4–27.

ENGSTED, T. AND S. V. MOLLER (2010): "An iterated GMM procedure for estimating the campbell-cochrane hait formation model, with an application to Danish stock and bond returns," *International Journal of Finance and Economics*, 15, 213–227.

EVENSEN, G. (1994): "Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical Research*, 99, 143–162.

FERNANDEZ-VILLAVERDE, J., J. F. RUBIO-RAMIREZ, T. J. SARGENT, AND M. W. WATSON (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97, 1021–1026.

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

HERBST, E. P. AND F. SCHORFHEIDE (2016): *Bayesian Estimation of DSGE Models*, Princeton University Press, 2016.

HOUTEKAMER, P.; MITCHELL, H. L. (1998): "Data assimilation using an ensemble Kalman filter technique," *Monthly Weather Review*, 126, 796–811.

JUDD, K. L. (1996): "Chapter 12 Approximation, perturbation, and projection methods in economic analysis," *Handbook of Computational Economics*.

KIM, J., S. KIM, E. SCHAUMBURG, AND C. A. SIMS (2008): "Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models," *Journal of Economic Dynamics and Control*, 32, 3397–3414.

KIM, J. AND S. H. KIM (2003): "Spurious welfare reversals in international business cycle models," *Journal of International Economics*, 60, 471–500.

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

KORDA, M. AND I. MEZIC (2017): "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator," *arXiv*.

MEZIC, I. (2022): "On Numerical Approximations of the Koopman Operator," *Mathematics 2022*, 10.

SCHMITT-GROHE, S. AND M. URIBE (2002): "Solving dynamic general equilibrium models using a second-order approximation to the policy function," Tech. rep., rutgers University and University of Pennsylvania.

TALLARINI, T. D. AND H. H. ZHANG (2005): "External Habit and the Cyclicality of Expected Stock Returns," *The Journal of business*, 78, 1023–1048.

WACHTER, J. A. (2005): "Solving Models With External Habit," *Finance Research Letters*, 2, 210–226.

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.