Structural estimation of rational expectations models with recursive preferences*

Bart F.C. Claassen — University of Groningen

Diego Ronchetti — Audencia Business School

Abstract

We propose a structural estimation method for the risk aversion and the ability to cope with uncertainty over time of a representative agent in a dynamic market equilibrium. The technique exploits the conditional restrictions for nonparametric state variables dynamics that describe the utility model in a Markovian setting. These restrictions are functional equations solved by Approximate Nonparametric Dynamic Programming based on a contraction mapping argument. Therefore, our method can accommodate any preference specification that features contraction, such as those in the Chew-Dekel class, and characterizes the corresponding parameter space. We study a representative agent endowed with Epstein-Zin utility for the U.S. equity and T-bill markets from 1952 to 2019. Our estimates of the preference parameters are plausible values that have not yet been established empirically. We include asymptotic properties of the model parameter estimators and evaluate the finite sample performance using a Monte Carlo study of the Bansal-Kiku-Yaron (2012) model.

JEL: C14, C36, E44, G12.

Keywords: structural estimation, contraction mapping, Local GMM, Approximate Nonparametric Dynamic Programming, Epstein-Zin preferences.

The need for a prompt resolution of risk or allowing its postponement comes from the endurance to cope with unpredictable scenarios over time. Research in

*We thank Tom Boot, Danilo Cavapozzi, Lammertjan Dam, Travis Johnson, Kenneth Judd, Walt Pohl, Eric Renault, Jules Tinang, and the participants at the MFA 2020 Annual Meeting, the SoFiE 2021 Annual Conference, the 2021 Empirical Macro-Finance Workcamp, the 2022 SOM Ph.D. Conference, the CFE 2022 International Conference, and seminars at Ca'Foscari University of Venice, and the University of Groningen for valuable comments.

economics and finance has shown that human decision-makers have preferences over the timing in the temporal resolution of uncertainty (e.g., Kreps and Porteus, 1978; Chew and Epstein, 1989; Dekel, 1986). Plausible characterizations of these preferences are necessary to calibrate models for analyzing (macro)economic shocks that, for example, policymakers and climate change induce and have long-run effects (e.g., Mumtaz and Theodoridis, 2020; Barnett et al., 2020). The time-non-separable preferences in the Chew-Dekel class, which nest the widespread Constant Relative Risk Aversion (CRRA) preferences, are popular specifications in utility models (e.g., Kreps and Porteus, 1978; Routledge and Zin, 2010; Backus et al., 2004; Campanale et al., 2010). Under these specifications, the preference over the early (or late) risk resolution is defined recursively over current consumption and a risk-adjusted valuation of the continuation utility. Specifically, a nonlinear forward-looking difference equation with a terminal condition defines the *next period*'s continuation utility.

We build an estimation method for Markovian nonparametric time-series models that are represented by functional equations that can be solved by a contraction mapping argument, such as the utility models discussed above. Our minimum-distance estimator is akin to a Local Generalized Method of Moments (LGMM), but with generalized residual functions whose functional forms are known just partially.[1] They are indeed known up to (**i**) an unknown Euclidean (finite-dimensional) parameter, and (**ii**) an unknown functional of this Euclidean parameter and the transition density of the state variables, which is itself unknown. Differently than in other econometric approaches, we reconstruct the unknown moment functions and parameter space accounting for their structural interdependence. In the first step, we rebuild the true conditional moment restrictions from a sequence of auxiliary ones based on a numerical convergence criterion. We reconstruct the moment function using Approximate Nonparametric Dynamic Programming (ANDP). For dynamic equilibrium models, the moment function includes the agent's continuation utility. We directly obtain it alongside its

[1]The LGMM extends the GMM in Hansen (1982) and Hansen and Singleton (1982) to directly account for conditional moment restrictions.

equivalent certainty measure, which we both need to compute the Stochastic Discount Factor (SDF). In the second step, we finalize the model estimation by exploiting the reconstructed conditional moment restrictions through an LGMM estimation for the nonparametric time-series model.[2]

Our method exploits *wholly* and *exclusively* the structural information that is embedded in the original model with a non-parametric specification of the Markovian dynamics of the state variables. This methodological improvement leads to higher precision in the estimation of the model parameters. For the utility models mentioned above, we use the information content of the dynamic pricing errors for the asset pricing model implied by the preference specification. We leave the structure of the SDF intact and exploit the flexibility of nonparametric regression techniques to reconstruct its value over time. We identify the continuation utility since any admissible lifetime utility function is a contraction mapping. In particular, the state variable dynamics and preference parameter values determine the *unique* fixed point for the continuation utility. For some combinations of recursive preferences and state variables dynamics, the existing literature provides limits to the parameter space (e.g., Hansen and Scheinkman, 2012; Borovička and Stachurski, 2020; Christensen, 2022). In other cases, our data-driven method allows determining the limits up to an error bound at the econometrician's discretion. This flexibility is because our approach does not rely upon (**i**) any ad-hoc parameterization of the dynamics of consumption growth and asset returns, (**ii**) any further assumption on the continuation utility, including proxying the wealth portfolio or hidden state variables, (**iii**) any relaxation of the interpretation of the SDF as a non-negative intertemporal marginal rate of substitution, and (**iv**) any instrument selection.

We apply our method to the estimation of the preference parameters of the representative U.S. consumer-investor from 1952Q1 to 2019Q3, adopting the homothetic preference specification introduced by Epstein and Zin (1989), which is

[2]Several authors stress the importance of retaining the entire information arising from the dynamic restrictions in economic and financial models (e.g., Domínguez and Lobato, 2004; Gospodinov and Otsu, 2012).

the most widely adopted among those in the Chew-Dekel class for economics and finance applications (e.g., Epstein and Zin, 1991; Chen et al., 2013; Brown and Kim, 2014). We summarize the U.S. equity return variation through the six size- and value-based bi-variate sorted FF portfolios of U.S. publicly traded equities and the 3-month T-Bills. We proxy consumption growth by data on Personal Consumption Expenditures of non-durables and services. Our point estimates correspond to values that are plausible from a theoretical perspective. They suggest the preference for an early resolution of risk, with a much lower upper bound of the confidence interval for the risk aversion parameter than previously found in the literature. Consequently, the equity premium puzzle becomes less severe. Confidence intervals obtained through existing estimation methods do not sustain this economic interpretation. We then argue that the sensitivity to structural information losses and model misspecification is considerable for estimating models with agents with recursive preferences.

1 Model structure

Time is discrete, labeled by $t \in \mathbb{Z}$. We denote the space of $(d^{[X]} \times d^{[X]})$-dimensional transition densities by $\mathcal{F}$, and a generic element of this space by f.

Assumption 1. *The random vector* $\boldsymbol{X}_t \in \mathcal{X} \subset \mathbb{R}^{d^{[X]}}$, *with* $d^{[X]} < \infty$ *summarizes the relevant information for pricing, and it is Markov of order one.*

The vector $\boldsymbol{X}_t$ may include exogenous factors for consumption growth and its volatility, dividend-price ratio, and term spreads. We denote its true one-period transition density and marginal probability density function by $f_\star \in \mathcal{F}$ and $f_\star^{[M]}$, respectively. If a variable is measurable w.r.t. the information available to investors, we say it is $\boldsymbol{X}_t$-measurable.

A unique non-durable consumption good serves as the numéraire.

Assumption 2. *A representative agent for the market consumes the amount of consumption good indicated by the function* $C : \mathcal{X} \mapsto \mathbb{R}_+$.

Assumption 2 coincides with point b in Assumption 1 of Hansen and Scheinkman (2012). As they explain, the function C is a convenient consumption specification that is determined endogenously.

1.1 The SDF family

The symbol $L_1(\mathcal{X})$ is for the linear space of integrable functions on $\mathcal{X}$. We use the symbol $\times$ for matrix dimensions and Cartesian product spaces. Consider a $d^{[\boldsymbol{\theta}]}$-dimensional real parameter vector $\boldsymbol{\theta}$ that we refer to as the SDF parameter vector, with $d^{[\boldsymbol{\theta}]} < \infty$, and an infinite-dimensional real parameter f serving as transition density for a $d^{[X]}$-dimensional real vector. The elements of $\boldsymbol{\theta}$ are the preference parameters.

The utility function $V_f(\boldsymbol{X}_t;\boldsymbol{\theta})$ represents the representative agent's (stationary) preferences over consumption strategies as follows. Recursive preferences in the Chew-Dekel class are defined by the time aggregator and the risk aggregator. The time aggregator F is a function that evaluates deterministic sequences of consumption. Koopmans (1960) characterizes the time aggregators that satisfy the conditions of historical independence, future independence, and stationarity. The risk aggregator $\mathcal{R}_{f;\boldsymbol{\theta}}$ measures the consequences and probabilities of risky pay-offs. Chew and Epstein (1989) and Dekel (1986) characterize a class of risk preferences that alleviates the independence axiom and include expected utility as a particular case. Backus et al. (2004) provide a coherent overview of this literature. The time aggregator and the risk aggregator jointly define the functional $V_{\cdot} : \mathcal{X} \times \mathcal{F} \times \mathbb{R}^{d^{[\boldsymbol{\theta}]}} \mapsto \mathbb{R}_+$ as

$$V_f(\boldsymbol{X}_t;\boldsymbol{\theta}) = F\Big(C(\boldsymbol{X}_t), \mathcal{R}_{f;\boldsymbol{\theta}}\left[V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})\right](\boldsymbol{X}_t);\boldsymbol{\theta}\Big), \tag{1}$$

where, in particular, the risk aggregator $\mathcal{R}_{f;\boldsymbol{\theta}}$, applied to the function $V_f(\cdot;\boldsymbol{\theta}) \in L_1(\mathcal{X})$ and valued at $\boldsymbol{X}_t$, returns the function $\mathcal{R}_{f;\boldsymbol{\theta}}\left[V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})\right](\boldsymbol{X}_t) \in L_1(\mathcal{X})$. The function $V_{f_\star}(\cdot;\boldsymbol{\theta})$ for the transition density $f_\star$ in Assumption 1 is the true continuation function.

Equation (1) constrains the SDF parameter vector's values and the state variables' dynamics. To emphasize the link between them, we introduce the conditional SDF parameter space for the transition density f as the space $\Theta_f \subseteq \mathbb{R}^{d^{[\boldsymbol{\theta}]}}$. The notation makes explicit that we focus on a subspace of $\mathbb{R}^{d^{[\boldsymbol{\theta}]}}$ identified by the fixed transition density f. The space Θ_f collects all the values $\boldsymbol{\theta}$ such that the function $V_f(\cdot;\boldsymbol{\theta})$ is a contraction and satisfies Equation (1). Therefore, for given dynamics of the state variables, this space includes all the allowable values that the SDF parameter vector can take while satisfying Equation (1). This definition embeds our identification strategy of the continuation value, which is global and nonparametric. It considers just the functional V. and, in the Epstein-Zin specification, it involves all the integrable functions.

Despite that our discussion concerns any preference specification that exhibits contraction, we particularize it to the case of Epstein-Zin preferences, for which $d^{[\boldsymbol{\theta}]} = 3$ and Equation (1) adapts to

$$V_f(\boldsymbol{X}_t;\boldsymbol{\theta}) = \left((1-\beta)C(\boldsymbol{X}_t)^{1-\frac{1}{\psi}} + \beta\,\mathrm{E}_f\left[V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})^{1-\gamma}\middle|\,\boldsymbol{X}_t\right]^{\frac{1}{\alpha}}\right)^{\frac{1}{1-\frac{1}{\psi}}}, \quad (2)$$

where $\mathrm{E}_f[\cdot|\,\boldsymbol{X}_t]$ is the conditional expectation operator for the transition density f by, the parameter $\beta \in (0:1)$ is a subjective discount factor, the parameter $\gamma > 0$ indicates the relative risk aversion, the parameter $\psi > 0$ is for the Elasticity of Intertemporal Substitution (EIS), and where we simplify the notation by the auxiliary parameter $\alpha := (1-\gamma)/(1-1/\psi)$. Therefore, with Epstein-Zin preferences, we are interested in the combinations of transition densities for a $d^{[X]}$-dimensional real vector and vectors in $\mathbb{R}^3_+$ satisfying Equation (2). We then analyze this equation for any transition density $f \in \mathcal{F}$ and any value for the parameter vector $\boldsymbol{\theta} = [\beta\ \gamma\ \psi]' \in \Theta_f \subseteq (0:1) \times \mathbb{R}^2_+$. The quantity $\mathrm{E}_f\left[V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})^{1-\gamma}\middle|\,\boldsymbol{X}_t\right]^{1/(1-\gamma)}$ is the certainty equivalence measure of $V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})$, which is concave because $\gamma > 0$. Consequently, the risk aggregator is lower when $V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})$ is more volatile. When $\gamma > 1/\psi$, the preferences imply a preference for the *early resolution of risk* (Kreps and Porteus, 1978; Bansal et al., 2010).

The SDF follows from the Benveniste-Scheinkman Theorem. Te first-order optimality condition of the consumption profile implies that the SDF

$$m_f(\boldsymbol{X}_{t+1},\boldsymbol{X}_t;\boldsymbol{\theta}) = \beta\frac{\left(V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})\,/\,\mathrm{E}_f\left[V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})^{1-\gamma}\middle|\,\boldsymbol{X}_t\right]^{\frac{1}{1-\gamma}}\right)^{-\left(\gamma-\frac{1}{\psi}\right)}}{\left(C(\boldsymbol{X}_{t+1})\,/C(\boldsymbol{X}_t)\right)^{\frac{1}{\psi}}}. \quad (3)$$

Consequently, the SDF variation over time is due to (**i**) the consumption growth rate $C(\boldsymbol{X}_{t+1})/C(\boldsymbol{X}_t)$; (**ii**) the unobservable continuation utility $V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})$; and (**iii**) the information spanned by the vector $\boldsymbol{X}_t$.[3]

The highly nonlinear relationships between state variables make the model estimation challenging. The historical consumption growth rate is close to independent and identically distributed (i.i.d.) and not that volatile around the value of 1. Intuitively, for a plausible value of the quarterly real consumption growth rate of 1.001, the values 1.001^2 and $1.001^{2/3}$ differ by roughly 10^{-3}, and the corresponding EIS parameter ψ is $1/2 = 0.5$ and $3/2 = 1.5$, respectively. Notwithstanding, these two distinct values have wildly different asset pricing implications and economic substance. When $\psi > 1$, the substitution effect dominates the wealth effect and vice versa. That is to say, when $\psi > 1$, agents buy more risky assets in response to higher expected growth, increasing the consumption-wealth ratio (e.g., see Bansal and Yaron, 2004). Furthermore, when $\gamma > 1/\psi$, investors prefer the early resolution of risk (e.g., see Backus et al., 2004). Next to that, when the consumption growth rate is i.i.d., the measure of certainty equivalence $\mathrm{E}\left[V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})^{1-\gamma}\middle|\,\boldsymbol{X}_t\right]^{1/(1-\gamma)}$ does not vary over time and, therefore, is rendered meaningless.[4] All in all, the modification of the utility from the CRRA case stems from just the scalar parameter ψ. However, the relationships between the relevant model variables are highly nonlinear, with risk and timing premia being structurally intertwined (Pohl et al., 2018).

To resolve the lack of observability of the continuation utility, Epstein and Zin (1991), Weber (2000), and Yogo (2006), among others, use the equivalent representation

$$M_{t+1} = \beta^{\alpha}\left(C(\boldsymbol{X}_{t+1})/C(\boldsymbol{X}_t)\right)^{-\frac{\alpha}{\psi}}\left(R_{t+1}^{[A]}\right)^{\alpha-1}, \tag{4}$$

where $R_{t+1}^{[A]}$ is the gross return from time t to time $t+1$ on the aggregate

[3]A CRRA utility function implies the SDF specification $m_f(\boldsymbol{X}_{t+1},\boldsymbol{X}_t;\boldsymbol{\theta}) = \beta\left(C(\boldsymbol{X}_{t+1})/C(\boldsymbol{X}_t)\right)^{-\gamma}$, which corresponds to the specification in Equation (3) for $\gamma = 1/\psi$.

[4]We show in Proposition 1 that this term is meaningless when the consumption growth rate is i.i.d., as $V_f(\boldsymbol{X}_{t+1};\boldsymbol{\theta})/C(\boldsymbol{X}_{t+1})$ is constant. The term between parentheses on the RHS of Equation (3) equals unity for all t, and Epstein-Zin preferences collapse to a CRRA specification.

consumption claim. It corresponds to the return on the portfolio associated with the representative agent's optimal investment strategy. Adapting Roll's (1977) critique of the Capital Asset Pricing Model, statistical tests might reject the model because the chosen proxy for the wealth portfolio does not capture the dynamics of the wealth portfolio.[5] Because Lettau and Ludvigson (2001) estimate that about two-thirds of total wealth consists of human capital, one has to be very careful in selecting or designing a proxy. Some researchers specify the complete macroeconomic structure alongside the agent's preferences (e.g., Bansal et al., 2016; Constantinides and Ghosh, 2011). Though the purpose of these studies is not necessary to validate the Epstein-Zin preferences, any misspecification of the macroeconomic structure will impact the statistical analyses of the preference parameters. Albeit their primary focus is not the validation of the Epstein-Zin preference specification, any wrong assumption on the design of the macroeconomic system may affect the inference of the preference parameters.

Detecting which unconditional restrictions, if any, out of the infinitely many dynamic ones implied by the model, retain the original information is also challenging (cf. Domínguez and Lobato, 2004). Stock and Wright (2000) suggest confidence sets that are immune to weak instruments. Yogo (2004) derives valid confidence intervals of a linearized specification of the SDF. However, Manresa et al. (2017) show that such strategies do not perform well in small samples. Linearization works well when γ and ψ are close to one, corresponding to log utility preferences. Suppose the risk aversion parameter γ is between five and ten, as commonly used in calibrations. In that case, linearization techniques might not be appropriate. Kleibergen and Zhan (2020) generalize the Stock and Wright (2000) approach to include many risk factors in both linear and nonlinear specifications. They allow for joint tests on pricing errors. However, the confidence sets they propose remain rather wide. Besides, other researchers exploit the information contents of modifications of the original model with recursive preferences. Yogo (2004)

[5] Roll (1977) argues in the context of the Capital Asset Pricing Model that a model specification test involving a proxy on the wealth portfolio constitutes a joint test on the model specification and the proxy's adequacy. Set statistical uncertainty aside, the rejection of the model may be due to its overall misspecification, the adopted lousy proxy of the wealth portfolio, or both.

linearizes the time-aggregator. Chen et al. (2013) particularize and filter the dynamics of the continuation utility-to-consumption ratio.

1.2 Stationary variables

The consumption level is typically non-stationary, and so is the value function $V_f(\cdot;\boldsymbol{\theta})$, which challenges the application of our method. To ease its implementation, we write the model in terms of the log-consumption growth rate $g : \mathcal{X}^2 \to \mathbb{R}$, defined as

$$g\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t\right) \coloneqq \ln\left[C\left(\boldsymbol{X}_{t+1}\right)/C\left(\boldsymbol{X}_t\right)\right] \tag{5}$$

(cf. Hansen and Scheinkman, 2012; Chen et al., 2013; Christensen, 2017). For $\psi \neq 1$, we express the model through the dynamic restriction

$$\begin{aligned}&\mathcal{A}_{f,\boldsymbol{\theta}}\left[v_f\left(\cdot;\boldsymbol{\theta}\right)\right]\left(\boldsymbol{X}_t\right)\\&\quad\coloneqq \mathrm{E}_f\left[e^{(1-\gamma)g(\boldsymbol{X}_{t+1},\boldsymbol{X}_t)}v_f\left(\boldsymbol{X}_{t+1};\boldsymbol{\theta}\right)^{\alpha} - \left(\frac{v_f\left(\boldsymbol{X}_t;\boldsymbol{\theta}\right)-1+\beta}{\beta}\right)^{\alpha}\middle|\boldsymbol{X}_t\right] = 0,\end{aligned} \tag{6}$$

featuring the unknown $\boldsymbol{X}_t$-measurable function $v_{\cdot} : \mathcal{X} \times \mathcal{F} \times \Theta_f \to \mathbb{R}$ defined as

$$v_f\left(\boldsymbol{X}_t;\boldsymbol{\theta}\right) \coloneqq \left(V_f\left(\boldsymbol{X}_t;\boldsymbol{\theta}\right)/C\left(\boldsymbol{X}_t\right)\right)^{1-1/\psi}. \tag{7}$$

If $\psi = 1$, the discussion adjusts with parameter space $\tilde{\Theta}_f \subseteq (0:1) \times \mathbb{R}_+$, $\boldsymbol{X}_t$-measurable function $\tilde{v}_{\cdot} : \mathcal{X} \times \mathcal{F} \times \tilde{\Theta}_f \to \mathbb{R}$ defined as

$$\tilde{v}_f\left(\boldsymbol{X}_t;\beta,\gamma\right) \coloneqq \ln\left[V_f(\boldsymbol{X}_t;\beta,\gamma,1)/C(\boldsymbol{X}_t)\right], \tag{8}$$

and dynamic restriction

$$\begin{aligned}&\tilde{\mathcal{A}}_{f,\boldsymbol{\theta}}\left[v_f\left(\cdot;\beta,\gamma\right)\right]\left(\boldsymbol{X}_t\right)\\&\quad\coloneqq \frac{\beta}{1-\gamma}\ln\left[\mathrm{E}_f\left[e^{(1-\gamma)\left(\tilde{v}_f(\boldsymbol{X}_{t+1};\beta,\gamma)+g(\boldsymbol{X}_{t+1},\boldsymbol{X}_t)\right)}\middle|\boldsymbol{X}_t\right]\right] - \tilde{v}_f\left(\boldsymbol{X}_t;\beta,\gamma\right) = 0.\end{aligned} \tag{9}$$

We can rewrite the SDF in Equation (3) as

$$m_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}) = \begin{cases} \beta \dfrac{e^{-\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} v_f\left(\boldsymbol{X}_{t+1}; \boldsymbol{\theta}\right)^{\alpha-1}}{\mathrm{E}_f\left[e^{(1-\gamma) g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} v_f\left(\boldsymbol{X}_{t+1}; \boldsymbol{\theta}\right)^{\alpha} \middle| \boldsymbol{X}_t\right]^{1-\frac{1}{\alpha}}}, & \psi \neq 1, \\ \beta \dfrac{e^{-\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t) + (1-\gamma)\tilde{v}_f(\boldsymbol{X}_{t+1}; \beta, \gamma)}}{\mathrm{E}_f\left[e^{(1-\gamma)\left(\tilde{v}_f(\boldsymbol{X}_{t+1}; \beta, \gamma) + g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)\right)} \middle| \boldsymbol{X}_t\right]}, & \psi = 1. \end{cases} \tag{10}$$

Online Appendix F contains the proofs of these results.

The terms on the left-hand side (LHS) of Equations (6) and (9) are functionals of the unknown continuation value functional $v_\cdot$. An analogous consideration holds for alternative recursive preference specifications. In general, the sensitivity of the dynamic restriction to changes in the parameter $\boldsymbol{\theta}$ at the borders of the space Θ_f depends on the functional $v_\cdot$. Ultimately, the functional $v_\cdot$ determines how informative the dynamic restriction is about the space Θ_f. We can describe the sensitivity of the functional $\mathcal{A}_{f,\boldsymbol{\theta}}\left[v_f\left(\cdot; \boldsymbol{\theta}\right)\right](\boldsymbol{X}_t)$ through its Fréchet derivative at $v_\cdot$ in the perturbation direction $\Delta v_\cdot$. This functional derivative indicates the first-order variation of the functional $\mathcal{A}_{f,\boldsymbol{\theta}}\left[v_f\left(\cdot; \boldsymbol{\theta}\right)\right](\boldsymbol{X}_t)$ to the perturbation of the continuation value $v_\cdot$ to the function $v_\cdot + \Delta v_\cdot$. Keeping all the remaining model elements fixed, we expand the functional in the following Fréchet first-order expansion:

$$\begin{aligned} &\mathcal{A}_{f,\boldsymbol{\theta}}\left[v_f\left(\cdot; \boldsymbol{\theta}\right) + \Delta v_f\left(\cdot; \boldsymbol{\theta}\right)\right](\boldsymbol{X}_t) \\ &\qquad = \left\langle D\mathcal{A}_{f,\boldsymbol{\theta}}\left[v_f\left(\cdot; \boldsymbol{\theta}\right)\right](\boldsymbol{X}_t), \Delta v_f\left(\cdot; \boldsymbol{\theta}\right)\right\rangle + O\left(\left\|\Delta v_f\left(\cdot; \boldsymbol{\theta}\right)\right\|_\infty^2\right), \end{aligned}$$

where the scalar $\left\|\Delta v_f\left(\cdot; \boldsymbol{\theta}\right)\right\|_\infty^2$ is the supremum norm of $\Delta v_f\left(\cdot; \boldsymbol{\theta}\right)$.

Our local identification assumption for $v_\cdot$ is an assumption on the local convexity of the functional $\mathcal{A}_{f,\boldsymbol{\theta}}\left[v_f\left(\cdot; \boldsymbol{\theta}\right) + \Delta v_f\left(\cdot; \boldsymbol{\theta}\right)\right](\boldsymbol{X}_t)$ at the function $v_f\left(\cdot; \boldsymbol{\theta}\right)$. We assume a local stationary point based on the sensitivity mentioned above. As we are interested in the entire space Θ_f and that space only, our identification assumption for the functional $v_\cdot$ also identifies the space $\Theta_{f_\star}$.[6]

[6] Identification conditions have been proposed in the literature for different settings than the one considered here. In a continuous-time Markov environment, Hansen and Scheinkman (2009) show that a function $\varsigma : \mathcal{X} \mapsto \mathbb{R}_{++}$, with the value e^{η} being the eigenvalue of the Perron-Frobenius equation $\mathbb{T}\varsigma(\boldsymbol{X}_t) = e^{\eta}\varsigma(\boldsymbol{X}_{t+t})$, for the operator $\mathbb{T}$ defined as $\mathbb{T}v_{f_\star}(\boldsymbol{X}_t; \boldsymbol{\theta}) =$

Assumption 3. *For the transition density $f_\star$ in Assumption 1 and any $\boldsymbol{\theta} \in \Theta_{f_\star}$, the functional equation* $\langle D\mathcal{A}_{f_\star,\boldsymbol{\theta}}\left[v_{f_\star}(\cdot;\boldsymbol{\theta})\right](\boldsymbol{X}_t), \Delta v_{f_\star}(\cdot;\boldsymbol{\theta})\rangle + O\left(\|\Delta v_{f_\star}(\cdot;\boldsymbol{\theta})\|_\infty^2\right) = 0$ *has the equivalently null function $\Delta v_{f_\star}(\cdot;\boldsymbol{\theta}) \equiv 0$ as unique solution.*

Assumption 4. *There is a partition of the state variable vector $\boldsymbol{X}_t = \left[\boldsymbol{Y}_t' \; \tilde{\boldsymbol{Y}}_t'\right]'$ such that the log-consumption growth rate at time t does not depend on $\tilde{\boldsymbol{Y}}_t$.*

To account for Assumption 4, we adjust the notation for the log-consumption growth rate in Equation (5) to $g(\boldsymbol{Y}_{t+1}, \boldsymbol{X}_t)$. We first introduce a new random vector $\boldsymbol{Y}_{t,*}$ and a new constant vector $\tilde{\mathbf{y}}_*$ such that we can solve the two equations in closed form when $\left[\boldsymbol{Y}_{t,*}' \; \tilde{\mathbf{y}}_*'\right]'$ replaces the state variable vector $\boldsymbol{X}_t$. Then, we construct the sequence of vectors indexed by the non-negative integer i

$$\left[\boldsymbol{W}_t'(i) \; \tilde{\boldsymbol{W}}_t'(i)\right]' := \frac{1}{i+1}\left[\boldsymbol{Y}_{t,*}' \; \tilde{\mathbf{y}}_*'\right]' + \left(1 - \frac{1}{i+1}\right)\left[\boldsymbol{Y}_t' \; \tilde{\boldsymbol{Y}}_t'\right]'. \tag{11}$$

As the index i increases, the sequence represents an increased perturbation of the reference random vector $\left[\boldsymbol{Y}_{t,*}' \; \tilde{\mathbf{y}}_*'\right]'$. The vector $\left[\boldsymbol{W}_t'(i) \; \tilde{\boldsymbol{W}}_t'(i)\right]' \in \mathcal{W} \times \tilde{\mathcal{W}} = \mathcal{X}$ with transition density $f_{[i]}$ induces the conditional expectation $\mathrm{E}_{[i]}\left[\cdot \middle| \boldsymbol{W}_t, \tilde{\boldsymbol{W}}_t\right]$. At the start of the sequence, $\mathrm{E}_{[0]}\left[\cdot \middle| \boldsymbol{W}_t, \tilde{\boldsymbol{W}}_t\right] \equiv \mathrm{E}_{[0]}\left[\cdot \middle| \boldsymbol{Y}_{t,*}, \tilde{\mathbf{y}}_*\right]$ is the conditional expectation induced by the vector $\left[\boldsymbol{W}_t'(0) \; \tilde{\boldsymbol{W}}_t'(0)\right]' \equiv \left[\boldsymbol{Y}_{t,*}' \; \tilde{\mathbf{y}}_*'\right]'$. At the end of the sequence, as $i \to \infty$, we have that $\lim_{i\to\infty} \mathrm{E}_{[i]}\left[\cdot \middle| \boldsymbol{W}_t, \tilde{\boldsymbol{W}}_t\right] \equiv \mathrm{E}_{f_\star}\left[\cdot \middle| \boldsymbol{X}_t\right]$ is the true conditional expectation we are ultimately interested in. We also introduce, for any positive integer i, the two functions $v_{[i]} : \mathcal{X} \times \mathcal{F} \times (0:1) \times \mathbb{R}_+^2 \mapsto \mathbb{R}_+$ and $\tilde{v}_{[i]} : \mathcal{X} \times \mathcal{F} \times (0:1) \times \mathbb{R}_+ \mapsto \mathbb{R}_+$ solving Equations (6) and (9) under the i-th perturbation of the reference random vector $\left[\boldsymbol{Y}_{t,*}' \; \tilde{\mathbf{y}}_*'\right]'$, and the conditional expectations

$$\mu_{[i]}^{[g]}(\tilde{w};\gamma) := \mathrm{E}_{[i]}\left[e^{(1-\gamma)g\left(\boldsymbol{W}_{t+1}(i), \tilde{\boldsymbol{W}}_t(i)\right)} \middle| \tilde{\boldsymbol{W}}_t = \tilde{w}\right], \tag{12}$$

for any $\tilde{w} \in \tilde{\mathcal{W}}$.

Proposition 1. *For the random vectors $\boldsymbol{Y}_t, \boldsymbol{Y}_{t,*} \in \mathbb{R}^{d^{[Y]}}$ and $\tilde{\boldsymbol{Y}}_t, \tilde{\boldsymbol{y}}_* \in \mathbb{R}^{d^{[\tilde{Y}]}}$ in*

$\mathrm{E}_{f_\star}\left[e^{(1-\gamma)g(\boldsymbol{X}_{t+1},\boldsymbol{X}_t)} v_{f_\star}(\boldsymbol{X}_{t+1};\boldsymbol{\theta}) | \boldsymbol{X}_t\right]$ for any $\boldsymbol{\theta} = [\beta \; \gamma \; \psi]' \in \Theta_{f_\star}$, characterizes a contraction mapping. The sign and magnitude of η then define $\Theta_{f_\star}$. For example, when $\eta < 0$ and $\beta < 1$, the restriction $-\ln[\beta] > \frac{1-1/\psi}{1-\gamma}\eta$ defines the bounds of $\Theta_{f_\star}$. Hansen and Scheinkman (2009) extend their arguments to the existence and uniqueness of the value function $\tilde{v}$.

Equation (11)*,* $v_{[0]}(\tilde{\boldsymbol{y}}_*;\boldsymbol{\theta}) = (1-\beta)/\left(1-\beta\mu_{[0]}^{[g]}(\tilde{\boldsymbol{y}}_*;\gamma)^{\frac{1}{\alpha}}\right)$ *and* $\tilde{v}_{[0]}(\tilde{\boldsymbol{y}}_*;\beta,\gamma) = (\beta/((1-\beta)(1-\gamma)))\ln\left[\mu_{[0]}^{[g]}(\tilde{\boldsymbol{y}}_*;\gamma)\right]$ *if **(i)*** $\mathbf{Y}_{t+1}(i) \perp \tilde{\mathbf{Y}}_{t+1}(i)|\boldsymbol{W}_t, \tilde{\boldsymbol{W}}_t, i$*; **(ii)** the functions* v_f *and* $\tilde{v}_f$ *are continuous in* $\tilde{\boldsymbol{W}}_t$ *and do not depend on* $\boldsymbol{W}_t$*; and **(iii)** the function* $\mu_{[0]}^{[g]}$ *in Equation* (12) *is finite for any* $\tilde{w} \in \tilde{\mathcal{W}}$*.*

Proof. See Appendix A. □

We take the solutions to Equations (6) and (9) in the setting considered in Proposition 1 as the functions $v_{[0]}$ and $\tilde{v}_{[0]}$. Then, for any $i \geqslant 1$, any $\tilde{w} \in \tilde{\mathcal{W}}$, and any $\boldsymbol{\theta} \in (0:1) \times \mathbb{R}^2$, we introduce the following value function iterations:

$$v_{[i]}(\tilde{w};\boldsymbol{\theta}) := 1-\beta+\beta\mathrm{E}_{[i]}\left[e^{(1-\gamma)g\left(\boldsymbol{W}_{t+1}(i),\tilde{\boldsymbol{W}}_t(i)\right)}v_{[i-1]}\left(\tilde{\boldsymbol{W}}_{t+1}(i);\boldsymbol{\theta}\right)^{\alpha}\middle|\tilde{\boldsymbol{W}}_t=\tilde{w}\right]^{\frac{1}{\alpha}}, \tag{13}$$

and

$$\tilde{v}_{[i]}(\tilde{w};\beta,\gamma) := \frac{\beta\ln\left[\mathrm{E}_{[i]}\left[e^{(1-\gamma)\left(\tilde{v}_{[i-1]}\left(\tilde{\boldsymbol{W}}_{t+1}(i);\beta,\gamma\right)+g\left(\boldsymbol{W}_{t+1}(i),\tilde{\boldsymbol{W}}_t(i)\right)\right)}\middle|\tilde{\boldsymbol{W}}_t=\tilde{w}\right]\right]}{1-\gamma}. \tag{14}$$

Online Appendix G discusses normal and Laplace log-consumption growth rates with time-varying first moments as reference cases for Proposition 1.

1.3 The true SDF

There are q risky assets traded in the economy, and $\boldsymbol{R}_t = [R_{1,t} \ \ldots \ R_{q,t}]'$ is the q-dimensional vector of cum-dividend gross returns. We denote the q-dimensional null vector and vector of ones by $\mathbf{0}_q$ and $\boldsymbol{\imath}_q$, respectively. The functional $\boldsymbol{h}_{\cdot} : \mathcal{X}^2 \times \mathcal{F} \times \Theta_f \times \mathbb{R}_+^q \to \mathbb{R}^q$ defined as $\boldsymbol{h}_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1};\boldsymbol{\theta}) := m_f(\boldsymbol{X}_{t+1};\boldsymbol{X}_t;\boldsymbol{\theta})\boldsymbol{R}_{t+1} - \boldsymbol{\imath}_q$ is the pricing error vector functional, defined for any $f \in \mathcal{F}$ and $\boldsymbol{\theta} \in \Theta_f$. Its conditional expectation $\boldsymbol{e}_{\cdot} : \mathcal{X} \times \mathcal{F} \times \Theta_f \to \mathbb{R}^q$ such that $\boldsymbol{e}_f(\boldsymbol{X}_t;\boldsymbol{\theta}) := \mathrm{E}_f\left[\boldsymbol{h}_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1};\boldsymbol{\theta})\middle|\boldsymbol{X}_t\right]$ is the conditional expected pricing error vector functional. If an additional asset offers at time t the risk-free gross return $R_{t+1}^{[F]}$ from that time and $t+1$, we adjust the population pricing error vector to $\boldsymbol{h}_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \tilde{\boldsymbol{R}}_{t+1};\boldsymbol{\theta}) := m_f(\boldsymbol{X}_{t+1};\boldsymbol{X}_t;\boldsymbol{\theta})\tilde{\boldsymbol{R}}_{t+1} - \mathbf{e}_{q+1}$, for the $(q+1)$-dimensional extended vector of cum-dividend returns in excess of the risk-free one $\tilde{\boldsymbol{R}}_t := \left[\left(\boldsymbol{R}_t - R_t^{[F]}\boldsymbol{\imath}_q\right)' \ R_t^{[F]}\right]'$, and the unit vector $\mathbf{e}_{q+1}$ with one as the $q+1$'th component.

From here onward, we consider the inclusion of a risk-free asset. Thus $\boldsymbol{h}_f$ is a $(q+1)$-vector. The following assumption, which states the absence of arbitrage possibilities, introduces the true value $\boldsymbol{\theta}_\star$ of the SDF parameter vector.[7]

Assumption 5. *For the transition density $f_\star$ in Assumption 1, the value $\boldsymbol{\theta}_\star \in \Theta_{f_\star}$ such that $\boldsymbol{e}_{f_\star}(\boldsymbol{X}_t; \boldsymbol{\theta}_\star) = \mathbf{0}_q$ is unique.*

2 Model estimation

We consider a nonparametric specification for the state variables dynamics and reconstruct it through kernel regression techniques.[8]

Assumption 6. *The process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is strictly stationary and time-homogeneous, and we have a sample of T time-series observations during $[1:T]$.*

We consider a kernel function $K : \mathbb{R}^{d^{[X]}} \mapsto \mathbb{R}_+$ such that $\int_{\mathbb{R}^{d^{[X]}}} K(\boldsymbol{x}) d\boldsymbol{x} = 1$, $\int_{\mathbb{R}^{d^{[X]}}} \boldsymbol{x} K(\boldsymbol{x}) d\boldsymbol{x} = \mathbf{0}_{d^{[X]}}$ and $\int_{\mathbb{R}^{d^{[X]}}} \boldsymbol{x}\boldsymbol{x}' d\boldsymbol{x} = \mathbf{I}_{d^{[X]}\times d^{[X]}} \int_{\mathbb{R}^{d^{[X]}}} x_i^2 K(\boldsymbol{x}) d\boldsymbol{x}$ and is independent of the index i, for the $(d^{[X]} \times d^{[X]})$-dimensional identity matrix $\mathbf{I}_{d^{[X]}\times d^{[X]}}$. Furthermore, we define the symmetric and positive-definite $(d^{[X]} \times d^{[X]})$-dimensional bandwidth matrix $\boldsymbol{H}$ for $\mathcal{K}^{\mathbf{H}}(\boldsymbol{x}) := |\boldsymbol{H}|^{-1/2} K(\boldsymbol{H}^{-1/2}\boldsymbol{x})$.[9] The kernel weight $w_T(\boldsymbol{x}, \boldsymbol{X}_s; \boldsymbol{H})$ for the state variable vector $\boldsymbol{X}_s$ conditionally on the value $\boldsymbol{x}$ is the s'th element of the $(T-1)$-vector $\boldsymbol{e}_1' \left(\boldsymbol{X}^{[\boldsymbol{X}]\prime} \boldsymbol{W}^{[\boldsymbol{W}]} \boldsymbol{X}^{[\boldsymbol{X}]}\right)^{-1} \boldsymbol{X}^{[\boldsymbol{X}]\prime} \boldsymbol{W}^{[\boldsymbol{W}]}$. In this expression, $\boldsymbol{X}^{[\boldsymbol{X}]}$ indicates either (**i**) the $(T-1)$-dimensional vector $\boldsymbol{\imath}_{T-1}$ in the case of the local constant kernel regression estimator (a.k.a. Nadaraya-Watson regression estimator); or (**ii**) the $((T-1) \times (d^{[X]}+1))$-dimensional matrix with $[1 \ (\boldsymbol{X}_s - \boldsymbol{x})']$ as s'th row in case of the local linear kernel regression estimator, for any $s = 1, \ldots, T-1$. We also use the $((T-1) \times (T-1))$-dimensional matrix $\boldsymbol{W}^{[\boldsymbol{W}]} := \text{diag}\left[\mathcal{K}^{\mathbf{H}}(\boldsymbol{X}_1 - \boldsymbol{x}), \ldots, \mathcal{K}^{\mathbf{H}}(\boldsymbol{X}_{T-1} - \boldsymbol{x})\right]$, for the vector-to-matrix diagonalization operator diag $[\cdot]$.

We denote by $\hat{f}_T$ and $\mathrm{E}_{\hat{f}_T}\left[\boldsymbol{\cdot} \,\middle|\, \boldsymbol{X}_t; \boldsymbol{H}\right]$ the kernel estimator of the one-period

[7] A few scalar parameters lack strong identification in typical parametric specifications of the state variables dynamics, such as the one in Bansal and Yaron (2004). Differently, here we identify nonparametrically the transition density $f_\star$.

[8] A global regression estimator that fits a whole curve over the entire sample space would spread local errors faster. The mistake at a point would affect the estimated regression in its proximity and the level and shape of the whole estimated regression curve.

[9] See, for example, Wand and Jones (1994).

transition density function of the process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ and a kernel regression function with bandwidth matrix $\boldsymbol{H}$, respectively. The matrix $\boldsymbol{H}$ controls the weighting of the observations around $\mathbf{x}$. Various kernel functions satisfy the assumptions above. However, the performance of kernel regressions is determined by the selected bandwidth $\boldsymbol{H}$, while the selected kernel function is of secondary importance (e.g. see Wand and Jones, 1994). We use the kernel of a multivariate normal distribution, $K(\boldsymbol{x}) \propto e^{-\boldsymbol{x}'\boldsymbol{x}}$, because it is numerically more stable than other kernel functions, such as the spherical Epanechnikov kernel $K(\boldsymbol{x}) \propto (1-\boldsymbol{x}'\boldsymbol{x})\,\mathbb{1}_{\{\boldsymbol{x}'\boldsymbol{x}\leqslant 1\}}$, in a multivariate setting.

2.1 The reconstruction of the SDF family

We estimate the conditional expectations in Equations (13) and (14) by ANDP. For any chosen $\left(d^{[X]}\times d^{[X]}\right)$-dimensional bandwidth matrix $\boldsymbol{H}^{[1]}$, we compute each expectation operator $\mathrm{E}_{\hat{f}_T}\left[\bullet\middle|\,\boldsymbol{X}_t;\boldsymbol{H}^{[1]}\right]$ by a local constant kernel regression.[10] For any point $[\boldsymbol{x}'\ \boldsymbol{\theta}']'$ on a grid on $\mathcal{X}\times\hat{\Theta}_T$, we compute the estimator of the functions $v_{[0]}(\mathbf{x};\boldsymbol{\theta})$ and $\tilde{v}_{[0]}(\mathbf{x};\beta,\gamma)$, which we denote by $\hat{v}_{[0]}\left(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{H}^{[1]}\right)$ and $\hat{\tilde{v}}_{[0]}\left(\boldsymbol{x};\beta,\gamma,\boldsymbol{H}^{[1]}\right)$.[11] Then, for positive integer numbers i and any grid point, we compute the function[12]

$$\begin{aligned}\hat{v}_{[i]}\left(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{H}^{[1]}\right) &= 1-\beta \\ &+\beta\left(\sum_{s=1}^{T-1} w_T\left(\boldsymbol{x},\boldsymbol{X}_s;\boldsymbol{H}^{[1]}\right)e^{(1-\gamma)g(\boldsymbol{X}_{s+1},\boldsymbol{X}_s)}\hat{v}_{[i-1]}\left(\boldsymbol{X}_{s+1};\boldsymbol{\theta},\boldsymbol{H}^{[1]}\right)^{\alpha}\right)^{\frac{1}{\alpha}} \qquad (15)\end{aligned}$$

if $\psi\neq 1$. When $\psi=1$, we apply the iterative scheme

$$\begin{aligned}\hat{\tilde{v}}_{[i]}\left(\boldsymbol{x};\beta,\gamma,\boldsymbol{H}^{[1]}\right) &= \frac{\beta}{1-\gamma} \\ &\cdot\ln\left(\sum_{s=1}^{T-1} w_T\left(\boldsymbol{x},\boldsymbol{X}_s;\boldsymbol{H}^{[1]}\right)e^{(1-\gamma)\left(g(\boldsymbol{X}_{s+1},\boldsymbol{X}_s)+\hat{\tilde{v}}_{[i-1]}\left(\boldsymbol{X}_{s+1};\beta,\gamma,\boldsymbol{H}^{[1]}\right)\right)}\right). \qquad (16)\end{aligned}$$

[10] The kernel regression estimators are asymptotically equivalent to integrals w.r.t. the transition density $\hat{f}_T$.

[11] The local constant kernel regression estimator ensures that any numerical reconstruction of the continuation utility function is a non-negative function.

[12] Our method resembles the Value Function Iteration (VFI) method that is often applied to solve computational equilibrium models (cf. Judd, 1998) in that we employ the empirical transition density rather than the transition density that is implied by the theoretical model.

For the reconstruction of the SDF, we apply Scott's rule of thumb bandwidth $\boldsymbol{H}_\star^{[1]} = T^{-\frac{1}{4+d[X]}} \mathrm{V}_{\hat{f}_T^{[M]}}\left[\boldsymbol{X}_t\right]$, where the matrix $\mathrm{V}_{\hat{f}_T^{[M]}}\left[\boldsymbol{X}_t\right]$ is the unconditional sample variance-covariance matrix of the vector $\boldsymbol{X}_t$. We select the maximum number of admissible iterations M and an error bound ϵ. For any fixed $\boldsymbol{\theta} = [\beta\ \gamma\ \psi]'$ with $\psi \neq 1$, the positive integer number $N_T(\boldsymbol{\theta})$, which may depend on T and $\boldsymbol{\theta}$, returns the number of iterations so that we define the estimated parameter space $\hat{\Theta}_T$ as

$$\hat{\Theta}_T := \left\{ \boldsymbol{\theta} \in \mathbb{R}^3_+ : \lim_{i \to N_T(\boldsymbol{\theta}) \leqslant M} \sup_{\boldsymbol{x} \in \mathcal{X}} \left| \frac{\hat{v}_{[i]}\left(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{H}_\star^{[1]}\right) - \hat{v}_{[i-1]}\left(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{H}_\star^{[1]}\right)}{\hat{v}_{[i-1]}\left(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{H}_\star^{[1]}\right)} \right| \leqslant \epsilon \right\}. \tag{17}$$

If $\psi = 1$, we replace the function $\hat{v}_{[i]}$ with $\hat{\tilde{v}}_{[i]}$.[13] Then, our reconstruction of the SDF in Equation (10) for any value $\boldsymbol{\theta} \in \hat{\Theta}_T$, which we denote by $m_{\hat{f}_T}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}, \boldsymbol{H}^{[1]}\right)$ is, for $\psi \neq 1$,

$$\frac{\beta e^{-\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} \hat{v}_{[N_T(\boldsymbol{\theta})]}\left(\boldsymbol{X}_{t+1}; \boldsymbol{\theta}, \boldsymbol{H}^{[1]}\right)^{\alpha-1}}{\left(\sum_{s=1}^{T-1} w_T\left(\boldsymbol{X}_t, \boldsymbol{X}_s; \boldsymbol{H}^{[1]}\right) e^{(1-\gamma) g(\boldsymbol{X}_{s+1}, \boldsymbol{X}_s)} \hat{v}_{[N_T(\boldsymbol{\theta})-1]}\left(\boldsymbol{X}_{s+1}; \boldsymbol{\theta}, \boldsymbol{H}^{[1]}\right)^{\alpha}\right)^{1-\frac{1}{\alpha}}}. \tag{18}$$

For $\psi = 1$, it is

$$\frac{\beta e^{-\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} e^{(1-\gamma)\hat{\tilde{v}}_{[N_T(\beta,\gamma)]}\left(\boldsymbol{X}_{t+1}; \beta, \gamma, \boldsymbol{H}^{[1]}\right)}}{\sum_{s=1}^{T-1} w_T\left(\boldsymbol{X}_t, \boldsymbol{X}_s; \boldsymbol{H}^{[1]}\right) e^{(1-\gamma)\left(g(\boldsymbol{X}_{s+1}, \boldsymbol{X}_s) + \hat{\tilde{v}}_{[N_T(\beta,\gamma)-1]}\left(\boldsymbol{X}_{s+1}; \beta, \gamma, \boldsymbol{H}^{[1]}\right)\right)}}, \tag{19}$$

where, again, we indicate the bandwidth matrix $\boldsymbol{H}^{[1]}$ as the last argument of the function. We then check the validity of Assumption 3 numerically on data, excluding multiple alternative specifications for the functional $v_{\hat{f}_T} := \hat{v}_{[N_T(\boldsymbol{\theta})]}\left(\cdot; \cdot, \boldsymbol{H}^{[1]}\right)$. In case $\psi \neq 1$, we look for a perturbation direction $\Delta v_{\hat{f}_T} \not\equiv 0$ solving the functional equation $\left\langle D\mathcal{A}_{\hat{f}_T, \boldsymbol{\theta}}\left[v_{\hat{f}_T}(\cdot; \boldsymbol{\theta})\right](\boldsymbol{X}_t), \Delta v_{\hat{f}_T}(\cdot; \boldsymbol{\theta})\right\rangle + O\left(\left\|\Delta v_{\hat{f}_T}(\cdot; \boldsymbol{\theta})\right\|_\infty^2\right) = 0$ for any

[13]In our empirical applications, choosing $v_{[0],\hat{f}_T}$ and $\tilde{v}_{[0],\hat{f}_T}$ either constant or varying as in Proposition 1 leads to the same convergence limit $v_{[N],\hat{f}_T}$ and $\tilde{v}_{[N],\hat{f}_T}$. However, in the second case, we need a far lower number of iterations. In particular, the difference in the number of iterations to achieve convergence is enormous for higher values of γ and values of ψ in the vicinity of 1. Because of the curvature of the risk aggregator, the lower γ, the closer is $v_{[0]}$ to the mean of $v_{[N],\hat{f}_T}$.

value $\boldsymbol{\theta} \in \hat{\Theta}_T$. In particular, we consider local perturbations of the reconstructed continuation value function and also families of interpolating functions from sets of basis functions. We proceed similarly with the functional $\tilde{v}_{\hat{f}_T} := \hat{\tilde{v}}_{[N_T(\boldsymbol{\theta})]}\left(\cdot;\cdot,\cdot,\boldsymbol{H}^{[1]}\right)$ in case $\psi = 1$.

We report the asymptotic distribution of the SDF reconstructed as in Equations (18) and (19) for any fixed value of the SDF parameter $\boldsymbol{\theta}$, under the standard simplifying parametrization of the bandwidth matrix $\boldsymbol{H}^{[1]}$ through the scalar bandwidth $b_T^{[1]}$, and regularity assumptions usually adopted in economic applications listed in Appendix C. We denote the conditional variance-covariance matrix operator for the state variables transition density f by $\mathrm{V}_f\left[\cdot \middle| \boldsymbol{X}_t\right]$.

Proposition 2. *Under Assumptions 7-15 in Appendix C, we have* $\hat{\Theta}_T \xrightarrow{\mathbb{P}} \Theta_{f_\star}$ *and* $\sqrt{T b_T^{[1] d[X]}}\left(m_{\hat{f}_T}(\tilde{\boldsymbol{x}}, \boldsymbol{x}; \boldsymbol{\theta}, b_T^{[1]}) - m_{f_\star}(\tilde{\boldsymbol{x}}, \boldsymbol{x}; \boldsymbol{\theta})\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \boldsymbol{K}\left(\boldsymbol{0}_{d[X]}\right) \mathcal{V}(\tilde{\boldsymbol{x}}, \boldsymbol{x}; \boldsymbol{\theta})\right)$ *for any* $\tilde{\boldsymbol{x}}, \boldsymbol{x} \in \mathcal{X}$, $\boldsymbol{\theta} \in \Theta_{f_\star}$, *the function convolution* $\boldsymbol{K}(\boldsymbol{x}) := \int_{\mathcal{X}} K(\boldsymbol{u})\, K(\boldsymbol{u} - \boldsymbol{x})\, d\boldsymbol{u}$ *and*

$$\mathcal{V}(\tilde{\boldsymbol{x}}, \boldsymbol{x}; \boldsymbol{\theta}) := \begin{cases} \dfrac{e^{-\gamma g(\tilde{\boldsymbol{x}}, \boldsymbol{x})} v_{f_\star}(\tilde{\boldsymbol{x}}; \boldsymbol{\theta})^{\alpha-1}\, \mathrm{V}_{f_\star}\left[e^{-\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} v_{f_\star}(\boldsymbol{X}_{t+1}; \boldsymbol{\theta})^{\alpha-1} \middle| \boldsymbol{X}_t = \boldsymbol{x}\right]}{\mathrm{E}_{f_\star}\left[e^{-2\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} v_{f_\star}^{2(\alpha-1)}(\boldsymbol{X}_{t+1}; \boldsymbol{\theta}) \middle| \boldsymbol{X}_t = \boldsymbol{x}\right] f_\star^{[M]}(\boldsymbol{x})}, & \text{for } \psi \neq 1, \\[2ex] \dfrac{e^{-\gamma g(\tilde{\boldsymbol{x}}, \boldsymbol{x}) + (1-\gamma)\tilde{v}_{f_\star}(\tilde{\boldsymbol{x}}; \beta, \gamma)}\, \mathrm{V}_{f_\star}\left[e^{-\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t) + (1-\gamma)\tilde{v}_{f_\star}(\boldsymbol{X}_{t+1}; \beta, \gamma)} \middle| \boldsymbol{X}_t = \boldsymbol{x}\right]}{\mathrm{E}_{f_\star}\left[e^{-2\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t) + 2(1-\gamma)\tilde{v}_{f_\star}(\boldsymbol{X}_{t+1}; \beta, \gamma)} \middle| \boldsymbol{X}_t = \boldsymbol{x}\right] f_\star^{[M]}(\boldsymbol{x})}, & \text{for } \psi = 1. \end{cases}$$

Proof. See Appendix D.2. □

For any point on a grid on $\mathcal{X}^2 \times \hat{\Theta}_T$ that is interesting for our analysis, we estimate the value $\mathcal{V}(\tilde{\boldsymbol{x}}, \boldsymbol{x}; \boldsymbol{\theta})$ by plugging $v_{[N_T(\boldsymbol{\theta})]}$ and $\tilde{v}_{[N_T(\boldsymbol{\theta})]}$ in its formula in place of $v_{f_\star}$ and $\tilde{v}_{f_\star}$, respectively, and estimating conditional expectation and variance by local linear kernel estimators $\mathrm{E}_{\hat{f}_T}\left[\bullet \middle| \boldsymbol{X}_t; \boldsymbol{H}^{[1]}\right]$ and $\mathrm{V}_{\hat{f}_T}\left[\bullet \middle| \boldsymbol{X}_t; \boldsymbol{H}^{[1]}\right]$, respectively. We consider $m_{\hat{f}_T}\left(\tilde{\boldsymbol{x}}, \boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{H}^{[1]}\right) e^{c_{\mathfrak{a}/2} \hat{V}_T(\tilde{\boldsymbol{x}}, \boldsymbol{x}; \boldsymbol{\theta}) / \left(\sqrt{T} m_{\hat{f}_T}\left(\tilde{\boldsymbol{x}}, \boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{H}^{[1]}\right)\right)}$ as the lower bound for the confidence interval for $m_{f_\star}(\tilde{\boldsymbol{x}}, \boldsymbol{x}; \boldsymbol{\theta})$, at the asymptotic confidence level $\mathfrak{a}$, where $c_{\mathfrak{a}/2}$ is the $(\mathfrak{a}/2)$-quantile of the standard normal distribution. We obtain the corresponding upper bound by replacing the $(\mathfrak{a}/2)$-quantile with the $(1 - \mathfrak{a}/2)$-quantile, from Proposition 2 and the Delta Method.

2.2 The estimation of the true SDF

We estimate the true value $\boldsymbol{\theta}_\star$ of the SDF parameter vector by a kernel-based LGMM estimator.[14] For any value $\{\boldsymbol{x}, \boldsymbol{\theta}\}$ on a grid of points over $\mathcal{X} \times \hat{\Theta}_T$, we compute the SDF defined in Equations (18) and (19) as follows. For any choice of the $\left(d^{[X]} \times d^{[X]}\right)$-dimensional bandwidth matrices in the set $\mathcal{H}^{[2]} := \left\{\boldsymbol{H}_1^{[2]}, \ldots, \boldsymbol{H}_q^{[2]}\right\}$ and $\boldsymbol{H}^{[1]}$, we implement the expectation operator $\mathrm{E}_{\hat{f}_T}\left[\bullet\middle|\, \boldsymbol{X}_t; \mathcal{H}^{[2]}\right]$ as a local linear kernel regression function:

$$\begin{aligned} e_{j,\hat{f}_T}&\left(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{H}^{[1]}, \boldsymbol{H}_j^{[2]}\right) \\ &:= \sum_{s=1}^{T-1} w_T\left(\boldsymbol{x}, \boldsymbol{X}_s; \boldsymbol{H}_j^{[2]}\right) h_{j,\hat{f}_T}\left(\boldsymbol{X}_{s+1}, \boldsymbol{X}_s, R_{j,s+1}; \boldsymbol{\theta}, \boldsymbol{H}_j^{[1]}\right), \end{aligned} \tag{20}$$

where the function $h_{j,\hat{f}_T}\left(\boldsymbol{X}_{s+1}, \boldsymbol{X}_s, \boldsymbol{R}_{s+1}; \boldsymbol{\theta}, \boldsymbol{H}^{[1]}\right)$ estimates the time s population pricing error for asset j for the SDF defined in Equations (18) and (19). We introduce the indicator function $\mathbb{1}_{\bar{\mathcal{X}}}$, which excludes pricing errors conditionally on the chosen part $\bar{\mathcal{X}}$ of the state variable space $\mathcal{X}$, to control for boundary effects in the kernel regression.[15] Moreover, we use local linear kernel regressions for the conditional pricing errors. These estimators suffer much less from boundary effects than local constant kernel regressors for a highly non-linear regression function (e.g. see Wand and Jones, 1994). Therefore, we can use a more lax trimming condition than if we would use local constant kernel regressions for the pricing errors.

We select each bandwidth matrix $\boldsymbol{H}_{j,\star}^{[2]}$, $j = 1, \ldots, q+1$, by leave-one-out cross-validation:

$$\boldsymbol{H}_{j,\star}^{[2]} := \underset{\boldsymbol{H}_j}{\operatorname{argmin}} \sum_{t=1}^{T-1} \left[h_{j,\hat{f}_T}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta}, \boldsymbol{H}_\star^{[1]}\right) - e_{j,\hat{f}_T,-t}\left(\boldsymbol{X}_t; \boldsymbol{\theta}, \boldsymbol{H}_\star^{[1]}, \boldsymbol{H}_j\right)\right]^2,$$

where $e_{j,\hat{f}_T,-t}$ is the j'th element of the kernel local linear kernel regression estimator vector defined in Equation (20) with the observation $\boldsymbol{X}_t$ excluded,

[14] Christensen (2017) estimates the model under the restriction $\psi = 1$ by a sieve method.

[15] In kernel regressions, trimming factors control boundary effects (cf. Tripathi and Kitamura, 2003; Gagliardini and Ronchetti, 2020).

and the minimization is over the set of admissible bandwidth matrices.

The estimator of the true value $\boldsymbol{\theta}_\star$ of the SDF parameter vector is

$$\hat{\boldsymbol{\theta}}_T := \underset{\boldsymbol{\theta}\in\hat{\Theta}_T}{\operatorname{argmin}} \frac{1}{T-1}\sum_{t=1}^{T-1} \mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)\boldsymbol{e}_{\hat{f}_T}\left(\boldsymbol{X}_t;\boldsymbol{\theta},\boldsymbol{H}_\star^{[1]},\mathcal{H}_\star^{[2]}\right)' \hat{\boldsymbol{\Omega}}_T\left(\boldsymbol{X}_t\right)^{-1}\boldsymbol{e}_{\hat{f}_T}\left(\boldsymbol{X}_t;\boldsymbol{\theta},\boldsymbol{H}_\star^{[1]},\mathcal{H}_\star^{[2]}\right), \quad (21)$$

for the set $\mathcal{H}_\star^{[2]} := \left\{\boldsymbol{H}_{1,\star}^{[2]},\dots,\boldsymbol{H}_{q,\star}^{[2]}\right\}$, and the trimming indicator function $\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{x})$ is one when $\boldsymbol{x}$ belong to the compact subset $\bar{\mathcal{X}}$ of the support of the state variables, and zero otherwise, and the $((q+1)\times(q+1))$-dimensional matrix $\hat{\boldsymbol{\Omega}}_t$ is a consistent estimator of a weighting matrix $\boldsymbol{\Omega}_\star$ chosen based on ease of implementation or properties. The estimator $\hat{\boldsymbol{\theta}}_T$ minimizes the time-series average of the empirical counterpart of a quadratic form of the pricing error vector based on the matrix $\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}$. In particular, the matrix weighs each pricing error on a traded asset. The quadratic form varies over time, and the criterion in Equation (21) is zero only when this form is identically null. Appendix D shows that this criterion function uniquely identifies the SDF parameter vector.

Depending on the chosen estimator properties to control, there are various natural candidates for the weighting matrix $\boldsymbol{\Omega}_\star$. For instance, the identity matrix weighs all pricing errors equally in the cross-section and across time. The conditional covariance matrix of the pricing errors minimizes the asymptotic variance of the estimator $\hat{\boldsymbol{\theta}}_T$. When the weighting matrix involves conditional expectations, such as the conditional covariance of the pricing errors, we employ a local constant regression with the common bandwidth matrix $\boldsymbol{H}_{\boldsymbol{\Omega}}$ for each element of the weighting matrix. This approach is standard in the literature on nonparametric estimation methods (e.g., Yin et al., 2010; Nagel and Singleton, 2011). The local constant estimator with a scalar bandwidth matrix ensures that $\hat{\boldsymbol{\Omega}}_t$ is positive definite. In addition, estimates that rely upon a unique bandwidth are numerically more stable than those obtained by $((q+1)\times q)$-dimensional bandwidth matrices for each weighting matrix element. Following Yin et al. (2010) and Nagel and Singleton

(2011), we employ the matrix

$$\underset{\boldsymbol{H}_{\boldsymbol{\Omega}}}{\operatorname{argmin}} \sum_{t=1}^{T-1} \ln \left[|\hat{\boldsymbol{\Omega}}_{T,-t}(\boldsymbol{X}_t; \boldsymbol{\theta}, \boldsymbol{H}_{\boldsymbol{\Omega}})| \right]$$
$$+ \left[\boldsymbol{h}_{\hat{f}_T} \left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta}, \boldsymbol{H}^{[1]} \right) - \boldsymbol{e}_{\hat{f}_T} \left(\boldsymbol{X}_t; \boldsymbol{\theta}, \boldsymbol{H}_\star^{[1]}, \mathcal{H}_\star^{[2]} \right) \right]' \hat{\boldsymbol{\Omega}}_{T,-t}^{-1}(\boldsymbol{X}_t; \boldsymbol{\theta}, \boldsymbol{H}_{\boldsymbol{\Omega}})$$
$$\cdot \left[\boldsymbol{h}_{\hat{f}_T} \left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta}, \boldsymbol{H}^{[1]} \right) - \boldsymbol{e}_{\hat{f}_T} \left(\boldsymbol{X}_t; \boldsymbol{\theta}, \boldsymbol{H}_\star^{[1]}, \mathcal{H}_\star^{[2]} \right) \right],$$

where $\hat{\boldsymbol{\Omega}}_{T,-t}(\cdot; \boldsymbol{\theta}, \boldsymbol{H}_{\boldsymbol{\Omega}})$ is the estimate of $\boldsymbol{\Omega}_\star$ obtained by omitting the t-th observation, and $|\cdot|$ denotes the matrix determinant. When $\boldsymbol{\Omega}_{\star,t}$ is the conditional centered (or uncentered) second-moment matrix of the pricing error vector, we consider its kernel estimator and compute the estimator $\hat{\boldsymbol{\theta}}_T$ iteratively, as with a Local Continuously Updated Generalized Method of Moments.

Proposition 3. *Under Assumptions 7-16 in Appendix C, estimator $\hat{\boldsymbol{\theta}}_T$ is consistent and asymptotically normal with $\sqrt{T}$-rate of convergence, with minimal asymptotic variance when $\boldsymbol{\Omega}_\star(\boldsymbol{X}_t) = \mathrm{V}_{f_\star} \left[\boldsymbol{h}_{f_\star} \left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta}_\star \right) \middle| \, \boldsymbol{X}_t \right]$.*

Proof. See Appendix D.3. □

3 The U.S. market from 1952 to 2019

We estimate the Epstein-Zin preference parameters for a representative agent that invests in the U.S. equity and T-bills market from 1952Q1 to 2019Q3. We assume that the log-consumption growth rate and the consumption-wealth ratio span the information used for pricing. We simplify the notation for the former in Equation (5) to g_t, and denote the latter by cay_t. We take consumption data from the National Income and Products Accounts tables of the Bureau of Economic Analysis. We proxy aggregate consumption by the sum of the *Personal Consumption Expenditures: Nondurables* and *Personal Consumption Expenditures: Services*, which we scale by *Population*. We obtain the consumption-wealth ratio of Lettau and Ludvigson (2001) from Martin Lettau's website.[16]

Our test assets are the 3-month T-Bill and the six value-weighted Fama-French portfolios, which are two-way sorted for the size and book-to-market factors. These

[16] https://sites.google.com/view/martinlettau/data

portfolios capture most of the cross-sectional variation in equity returns (e.g., Fama and French, 1993; Nagel and Singleton, 2011). They classify as Growth (G), Neutral (N), and Value (V), along the book-to-market dimension and as Small (S) and Big (B), along the market capitalization dimension. Accordingly, we add the superscripts SG, SN, SV, BG, BN, and BV to the notation of the gross time t returns. The returns on these portfolios are from Kenneth French's website.[17] We proxy the risk-free rate $R_t^{[F]}$ by the gross return on 3-month T-Bills from Federal Reserve Economic Data (FRED). We adjust consumption and the asset returns for inflation by scaling the data for the *Personal Consumption Expenditures Chain-type Price Index* from the Bureau of Economic Analysis.

Figure 1 displays the log-consumption growth rate g_t and the consumption-wealth cay_t over time. The National Bureau of Economic Research (NBER) recession periods from the FRED recession indicators are the shaded areas. No volatility clustering and persistence are apparent in the log-consumption growth rate, which suggests that it is close to i.i.d.. We standardize both conditioning variables for numerical stability. We trim through the set $\mathcal{X}_\star = \left\{ \boldsymbol{X}_t \in \mathbb{R}^2 : \sqrt{X_{1,t}^2 + X_{2,t}^2} \leqslant 3 \right\}$. Figure 2 exhibits this trimming. We retain just the points inside the circle. In practice, we exclude very little data.

Since consumption growth and the consumption-wealth ratio predict conditional returns well, we assume they span the information set. Figure 3 exhibits the (excess) expected returns on the test assets conditional on the log-consumption growth rate and consumption-wealth ratio. We can see that both variables capture different components of (excess) returns, as evidenced by the gradients' diagonal directions.[18]

We compute the estimator $\hat{\Theta}_T$ selecting the combinations of values of the preference parameters that make the iterations in Equations (15) and (16) converge. Figure 4 exhibits the estimated parameter space. At each node in dark gray, $v_{\hat{f}_T}$ and $\tilde{v}_{\hat{f}_T}$ exhibit contraction. Panel 4(a) refers to the instance that we restrict ψ to unity. We construct an orthogonal grid with $0.99 \leqslant \beta \leqslant 1.002$ and $0.5 \leqslant \gamma \leqslant 150$

[17] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[18] This result is unsurprising because the two variables do not correlate (their estimated correlation equals 0.02), while both variables describe nonlinear relations with expected (excess) returns. Figures in Online Appendix E show this effect.

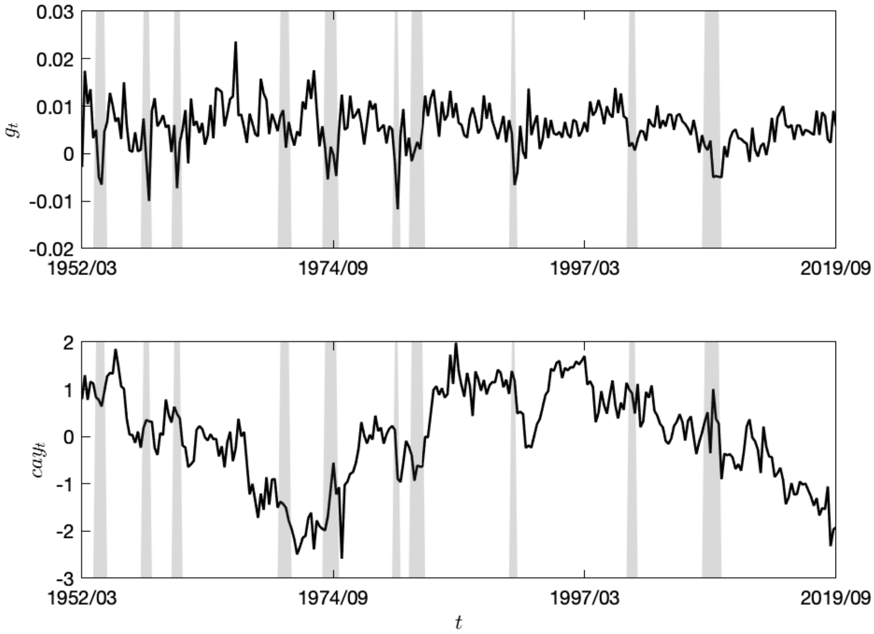


Figure 1. The log-consumption growth rate, g_t, and the standardized consumption-wealth ratio, cay_t, with null mean and unit variance, over time. The shaded areas correspond to the NBER recession periods.

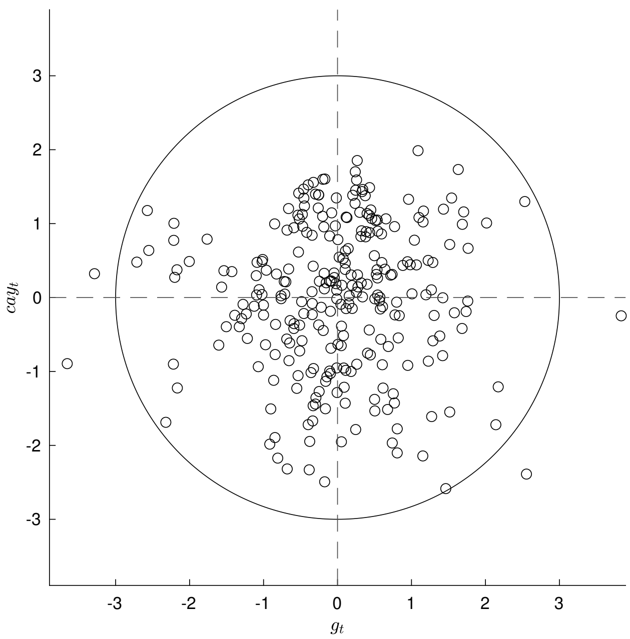


Figure 2. Trimming using a circle with a radius of 3 for the standardized values of log-consumption growth rate, g_t (on the horizontal axis), and the consumption-wealth ratio, cay_t (on the vertical axis).

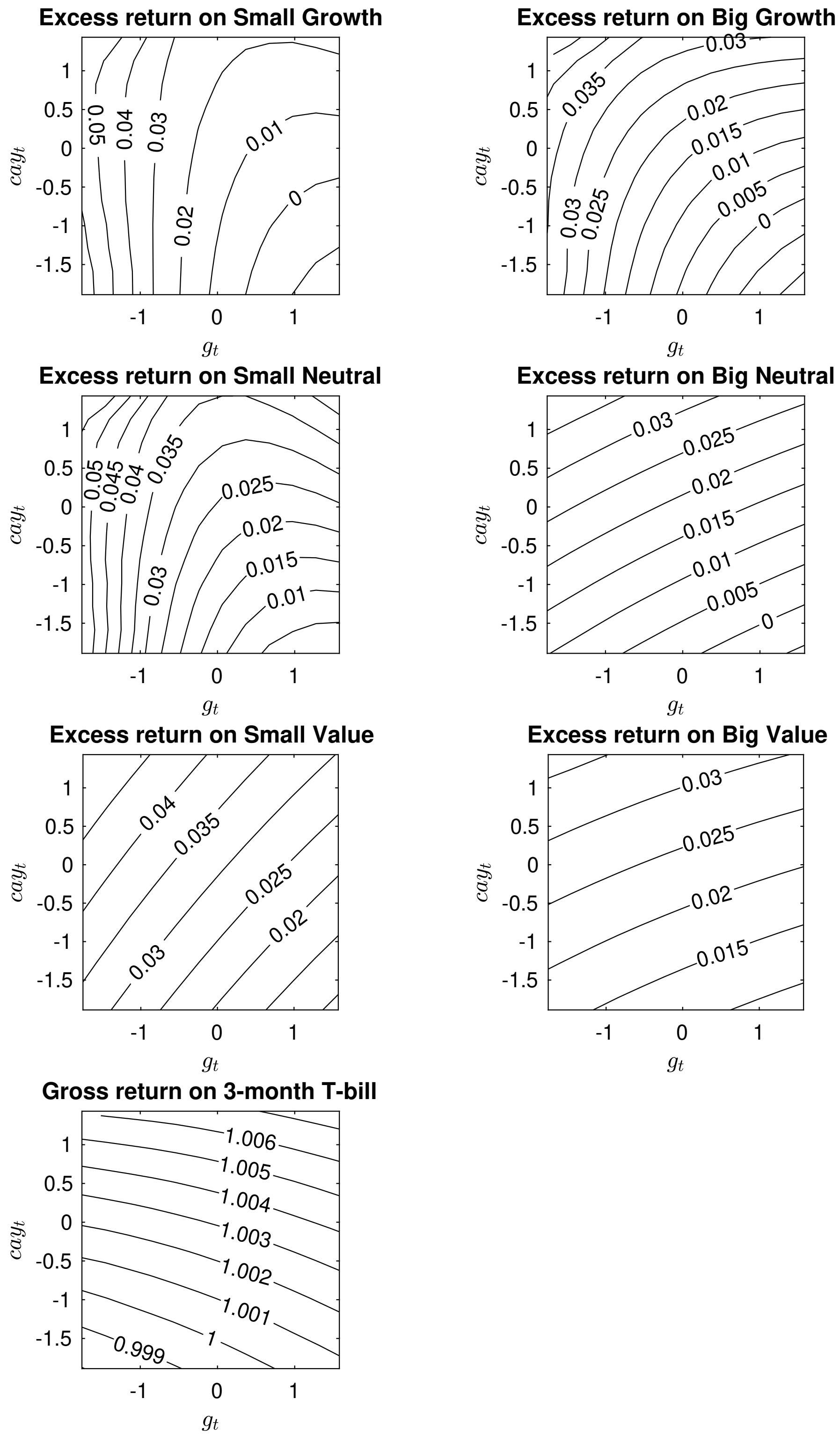
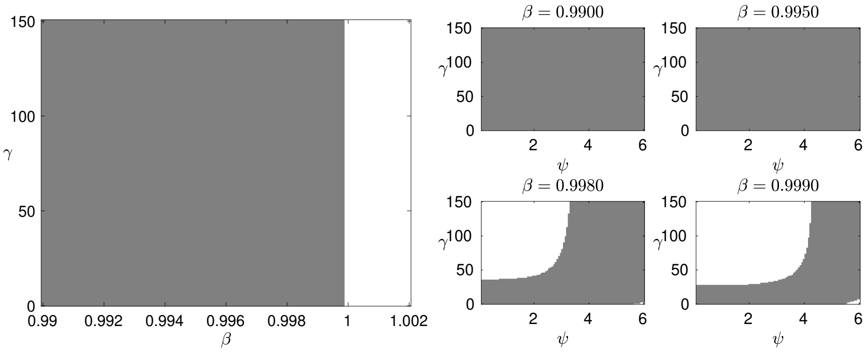


Figure 3. Conditional expected (excess) returns on both log-consumption growth rate, g_t (on the x-axes), and the consumption-wealth ratio, cay_t (on the y-axes). We standardize the values of g_t and cay_t to a mean of zero and a standard deviation of one for numerical stability.

and take 100 equidistant nodes in each direction. Thus, we evaluate $10,000$ points on the two-dimensional grid. We can see that we have a contraction for all values of γ between 0.5 and 150 and for $\beta \leqslant 0.9993$. Given that we use quarterly data, we think this is a very reasonable upper bound on β.



(a) $\psi = 1$. (b) $\psi \neq 1$.

Figure 4. Preference parameter space estimated by the method in Section 2 and log-consumption growth and consumption-wealth ratio spanning the information used to set prices.

Panel 4(b) refers to the case with $\psi \neq 1$. Since β is the primary determinant of the speed and limit of convergence among the preference parameters, we consider some of its values. For each of them, we construct an orthogonal grid with $0.5 \leqslant \gamma \leqslant 150$ and $0.1 \leqslant \psi \leqslant 6$ with 100 nodes in each direction. Given that we have quarterly data, we can expect an estimate of β in the vicinity of 0.993. Therefore, we construct grids for β equal to 0.99 and 0.995. In addition, we evaluate the convergence for relatively high values of β since they are associated with weaker convergence. Specifically, we use $\beta = 0.998$ and $\beta = 0.999$, which are relatively high values β since such values are associated with a monthly rather than a quarterly frequency. For $\beta = 0.99$ and $\beta = 0.995$, we have convergence everywhere. However, for values of β closer to 1, we see that the maximal γ varies with the value of ψ. When ψ is lower than 4, the maximum value of γ drops substantially. This fact is rather unsurprising since for a low value of ψ and a high value of γ the absolute value of the auxiliary parameter $\alpha := (1-\gamma)/(1-1/\psi)$ in Equation (15) becomes very high. However, such values for β are implausible for quarterly data, so the space $\hat{\Theta}_T$ is large and unlikely to constrain the other SDF parameters.[19]

Table 1 reports the point estimates and serial bootstrap confidence intervals of the preference parameters obtained with the identity matrix as the weighting matrix. This weighting matrix, which weighs the pricing errors on the distinct assets equally in the cross-section and across time, is the most interesting from an economic perspective. The conditional covariance matrix of the pricing errors minimizes the asymptotic variance of the SDF parameter vector estimator. Still, it puts a lot of weight on the risk-free asset since the variation of pricing errors is relatively low. The first column with numbers exhibits the estimates for the estimates of β and γ when we restrict ψ to the value of 1. The former is at 0.989, which is somewhat low given that we have quarterly data. At the same time, the estimate of γ is slightly higher than expected at a value of 16.55. The second column with numbers exhibits the estimates of the parameters when ψ is a free parameter. We can see that the point estimates of β and γ do not change that much but that the confidence interval of γ becomes substantially wider. The estimated value of ψ is 1.74 but is not statistically significantly bigger than 1.

Table 1. Estimates of the preference parameters and minimized criterion $\hat{\mathcal{Q}}_T$ that define the estimator $\left[\hat{\beta}_T \; \hat{\gamma}_T \; \hat{\psi}_T\right]'$ in Equation (21), with serial bootstrap 90% standard confidence intervals in parentheses (expected block length of six). The estimation method is described in Section 2, with the identity matrix as the weighting matrix, the six FF portfolios, and the 3-month T-Bill as test assets, and log-consumption growth rate and consumption-wealth ratio as conditioning information to set prices at time t. Online Appendix H describes the serial bootstrap.

$\hat{\beta}_T$	0.989	0.987
	(0.967, 0.999)	(0.972, 1.00)
$\hat{\gamma}_T$	16.55	16.44
	(2.50, 23.94)	(1.85, 40.27)
$\hat{\psi}_T$	-	1.74
		(0.10, 2.66)
$\hat{\mathcal{Q}}_T$	0.004	0.0039
	(0.002, 0.009)	(0.002, 0.009)

[19] We reconstruct the SDF family as described in Section 2 with normal and Laplace log-consumption growth rates as reference cases.

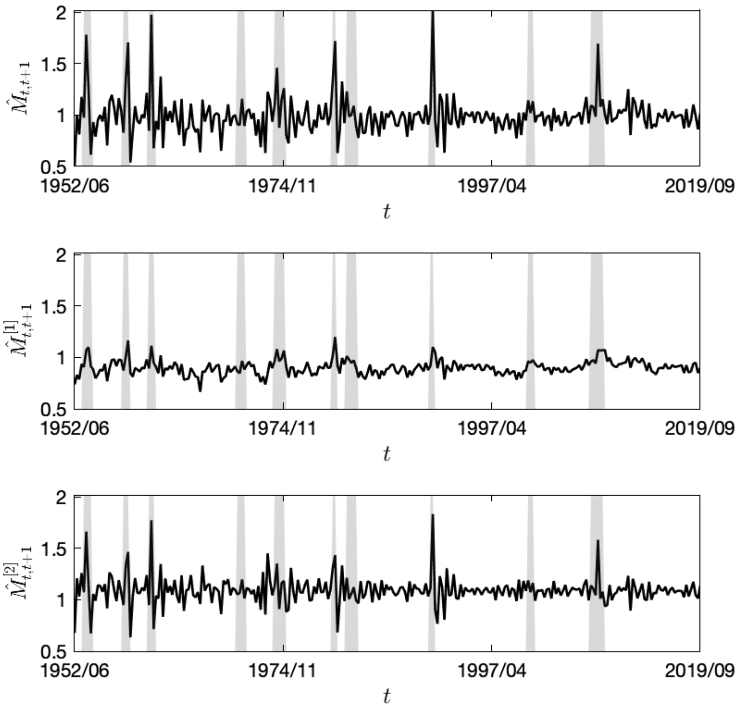


Figure 5. Estimated SDF $\hat{M}_{t,t+1}$ and its multiplicative components $\hat{M}^{[1]}_{t,t+1} := \hat{\beta}_T \left(C\left(\boldsymbol{X}_{t+1}\right)/C\left(\boldsymbol{X}_t\right)\right)^{-\frac{1}{\hat{\psi}_T}}$ and $\hat{M}^{[2]}_{t,t+1} := \hat{M}_{t,t+1}/\hat{M}^{[1]}_{t,t+1}$ over time.

The first panel of Figure 5 displays the estimated SDF as a function of time. It exhibits heteroskedasticity, and it is relatively volatile around recessions. The second and third panels of the figure show the SDF multiplicative components $\beta\left(C\left(\boldsymbol{X}_{t+1}\right)/C\left(\boldsymbol{X}_t\right)\right)^{-\frac{1}{\psi}}$ and $\left(\dfrac{V_f\left(\boldsymbol{X}_{t+1};\boldsymbol{\theta}\right)}{\mathrm{E}_f\left[V_f\left(\boldsymbol{X}_{t+1};\boldsymbol{\theta}\right)^{1-\gamma}\middle|\boldsymbol{X}_t\right]^{\frac{1}{1-\gamma}}}\right)^{-\left(\gamma-\frac{1}{\psi}\right)}$, respectively. In particular, the latter component introduces most of the heteroskedasticity. This pattern explains our relatively low estimate compared to estimates of γ in applications with the CRRA discount factor. With that preference specification, $\gamma = 1/\psi$, so that the latter SDF component is everywhere equal to 1. In that case, one has to assume a very high value for the risk aversion parameter γ, say 90, to generate sufficient variation in the SDF to match the equity premium (Mehra and Prescott, 1985).[20] With Epstein-Zin preferences, as the estimate of the second component of the SDF is somehow volatile, we do not need to inflate further the variation in consumption growth rate by raising it to a high power in the first SDF

[20]Mehra and Prescott (1985) suggest a lower bound of four and an upper bound of ten for γ.

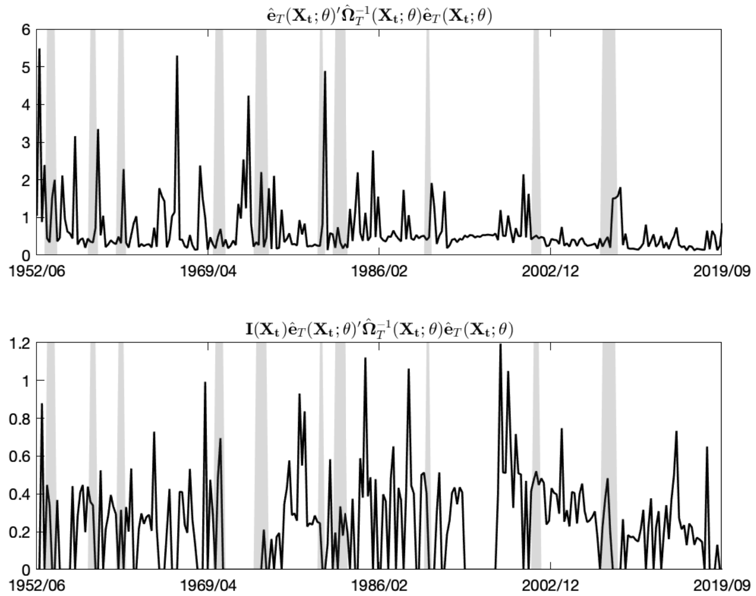


Figure 6. Quadratic form of the pricing errors of the six FF portfolios and the 3-month T-Bill at the estimate $\hat{\boldsymbol{\theta}}_T = [0.987\ 16.44\ 1.74]'$ over time. The estimate is for the six FF portfolios and the 3-month T-Bill as test assets, consumption growth and consumption-wealth ratio spanning the information used to set prices at time t, and the identity matrix as the weighting matrix.

component.

The top panel of Figure 6 plots the value $\hat{\boldsymbol{e}}(\boldsymbol{X}_t;\hat{\boldsymbol{\theta}}_T)'\hat{\boldsymbol{e}}(\boldsymbol{X}_t;\hat{\boldsymbol{\theta}}_T)$ as a function of time. The pricing errors are substantial during recessions and crises. The consumption growth rate takes on extremely low values during such periods, as shown in the upper panel of Figure 1. However, the indicator function nullifies such periods in the criterion function (21). The bottom panel of Figure 6 displays the statistic $\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)\hat{\boldsymbol{e}}(\boldsymbol{X}_t;\hat{\boldsymbol{\theta}}_T)'\hat{\boldsymbol{e}}(\boldsymbol{X}_t;\hat{\boldsymbol{\theta}}_T)$ as a function of time. The asset pricing model explains the data very well in more tranquil times. Next, we can understand that we might need to generate more volatility in the SDF to match the price fluctuations in turbulent periods by raising γ.

4 Monte Carlo experiment

We evaluate the finite sample performance of our estimator with a Monte Carlo (MC) experiment of the Bansal, Kiku and Yaron 2012 (BKY) long-run risks model. In the model for q tradable risky assets, the $(q+3)$-dimensional vector $\boldsymbol{\zeta}_t = [\eta_t \; u_{1,t} \; \dots \; u_{q,t} \; \varepsilon_t \; w_t]' \sim \mathcal{N}(\mathbf{0}_{q+3}, \boldsymbol{I}_{q+3})$ of unobservable shocks is at the base of the dynamics of the log-consumption growth rate, $g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)$, and of the growth rate of each dividend claim, $g_j(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)$ for $j = 1, \dots, q$. The function μ induces persistent changes in the expected growth rates of consumption and dividends. The function σ represents the stochastic volatility component that creates conditional heteroskedasticity in these growth rates. The following set of equations characterize the dynamics of the BKY economy:

$$\begin{cases} g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t) = \mu_0 + \mu(\boldsymbol{X}_t) + \sigma(\boldsymbol{X}_t)\eta_{t+1}, \\ g_j(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t) = \mu_j + \phi_j \mu(\boldsymbol{X}_t) + \sigma(\boldsymbol{X}_t)\left(\pi_j \eta_{t+1} + \varphi_j u_{j,t+1}\right), \\ \mu(\boldsymbol{X}_{t+1}) = \rho\mu(\boldsymbol{X}_t) + \varphi^{[e]}\sigma(\boldsymbol{X}_t)\varepsilon_{t+1}, \\ \sigma^2(\boldsymbol{X}_{t+1}) = \bar{\sigma}^2 + \nu\left(\sigma^2(\boldsymbol{X}_t) - \bar{\sigma}^2\right) + \sigma^{[w]} w_{t+1}, \end{cases} \tag{22}$$

where $|\rho| < 1$, $\varphi_j, \varphi^{[\epsilon]}, \sigma^{[w]} > 0$, $\mu_0, \mu_j \in \mathbb{R}$. Even if the shocks are independent, the variables $g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)$ and $g_j(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)$ are correlated because of the common factor $\mu(\boldsymbol{X}_t)$. In calibrations, the factor μ is persistent since ρ is close to one and $\sigma(\boldsymbol{X}_t)$ is relatively small (e.g., Bansal and Yaron, 2004; Bansal et al., 2012). Additionally, note that the log-consumption growth rate and dividend growth rate are i.i.d. when $\sigma^{[w]} = \varphi^{[\epsilon]} = 0$. The parameters $\phi_j > 1$ and $\pi_j > 1$ govern j-th dividend dynamics, particularly its mean, variance, and correlation with the log-consumption growth rate. For $q = 1$ and $\pi_1 = 0$, the model is to the Bansal and Yaron (2004) one.

We align our MC experiment with the empirical estimation by adopting a quarterly calibration and 275 simulated quarters. We solve the model using global projection methods. Therefore, we do not induce approximation errors that result from i) log-linearization or ii) errors from time aggregation of a high-frequency simulation to quarterly observations. Online Appendix I explains how we solve the model using global projection methods. We generate MC samples by simulating the returns using this global solution. Furthermore, we note that the functions $\mu(\boldsymbol{X}_t)$ and $\sigma^2(\boldsymbol{X}_t)$ contain the necessary information for the time t conditional expectations and can serve as the Markov state variables. Therefore, we proceed with our discussion with $\mathcal{X} = \mathbb{R} \times \mathbb{R}_+$ and the same notation as in Proposition 1.

4.1 Setting

We evaluate the estimator's finite sample properties in $B = 1000$ time series. We simulate the processes for six risky gross returns and a risk-free gross return. We take unconditional averages of the variables as starting values, and we use a burn-in period to mitigate the potential influence of our choice. Table 2 displays the DGP parameter values, which correspond with a quarterly calibration of the (Bansal et al., 2012) model.

Table 2. Calibration of the MC sample

Preferences		*Volatility*		*Consumption*	
β	0.9930	$\bar{\sigma}$	0.0165	μ_0	0.0045
γ	10	$\nu^{[1]}$	0.9960	ρ	0.9000
ψ	1 or 2	$\sigma^{[w]}$	0 or 0.89e-5	$\varphi^{[e]}$	0.0380

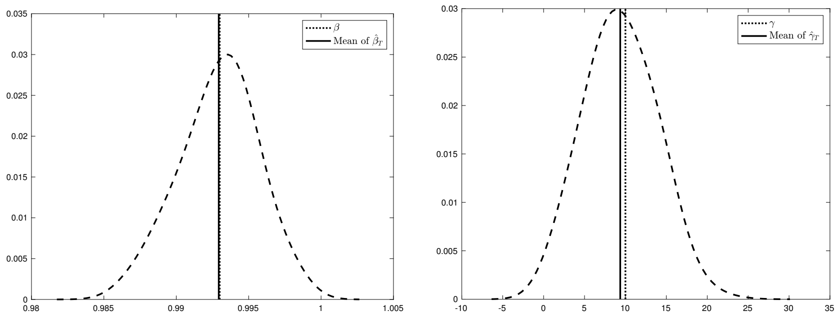
Asset	μ_j	ϕ_j	π_j	φ_j
1	0.0045	1.6000	2.0000	5.0000
2	0.0045	2.1000	2.6000	5.2500
3	0.0045	2.5000	2.8000	5.5000
4	0.0045	2.7500	2.9000	5.7500
5	0.0045	3.0000	2.9500	6.0000
6	0.0045	3.2000	3.1000	6.5000

First, we consider $\psi = 1$ and no stochastic volatility, that is $\sigma^{[w]} = 0$. We then estimate the parameters with ψ unconstrained and $\sigma^{[w]} = 0$. The simulation model parameters are the same as before, except for ψ, which we set equal to 2, as do Bansal et al. (2012).

4.2 Results

Figure 7 displays the empirical distribution of the estimated SDF parameters for $\psi = 1$ and $\sigma^{[w]} = 0$. Panel 7a exhibits the distribution of $\hat{\beta}_{b,T}^{[MC]}$. The dashed vertical line is at the Data Generating Process (DGP) parameter value of 0.993, and the solid vertical line indicates the mean of its $1,000$ bootstrap estimates. We do not find any bias in the estimation of β.

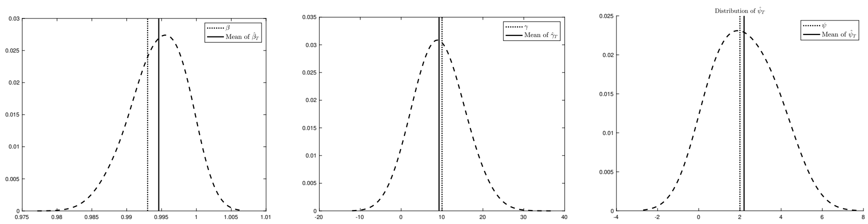
Panel 7b exhibits the distribution of $\hat{\gamma}_{b,T}^{[MC]}$. The dashed vertical line indicates the DGP value of 10 for the risk aversion parameter, and the solid vertical line is at



(a) Distribution of $\hat{\beta}_T$ (b) Distribution of $\hat{\gamma}_T$

Figure 7. Smoothed distribution of the MC estimates of β and γ in the BKY long-run risks model for six risky assets and a risk-free one, with $\psi = 1$ and log-consumption growth rate spanning the information used to set prices. In each MC iteration, we estimate the SDF parameter vector by the method described in Section 2, with the identity matrix as the weighting matrix.

the mean of its $1,000$ bootstrap estimates $\hat{\gamma}_{b,T}^{[MC]}$. We cannot point to a downward bias because of the magnitude of the MC standard errors.



(a) Distribution of $\hat{\beta}_T$ (b) Distribution of $\hat{\gamma}_T$ (c) Distribution of $\hat{\psi}_T$

Figure 8. Smoothed distribution of the MC estimates of β, γ, and ψ, in the BKY long-run risks model for six risky assets and a risk-free one, with log-consumption growth rate spanning the information used to set prices. In each MC iteration, we estimate the SDF parameter vector by the method described in Section 2, with the identity matrix as the weighting matrix.

Panel 8a exhibits the distribution of $\hat{\beta}_{b,T}^{[MC]}$. The dashed vertical line is at the DGP parameter value of 0.993, and the solid vertical line indicates the mean of its $1,000$ bootstrap estimates. We cannot point to an upward bias because of the magnitude of the MC standard errors. Also, we note that the distribution of this estimate is left-tailed. Panel 8b exhibits the distribution of the MC estimate $\hat{\gamma}_{b,T}^{[MC]}$. The dashed vertical line indicates the DGP value of 10 for the risk aversion parameter γ, and the solid vertical line indicates the mean of its bootstrap estimates.

In Panel 8c, we report the distribution of the MC estimates of ψ. The dashed line indicates the DGP value of 2 and we see that the mean of the MC estimates, as indicated by the solid line, is slightly to the right of this true value. By comparing Figures 7 and 8, we can also observe that the confidence intervals for β and γ are somewhat wider when ψ is a free parameter.

5 Conclusion

We propose a minimum distance estimation method for nonparametric time series models described by conditional moment restrictions as functional equations solved by a contraction mapping argument. The technique exploits the complete structural information of the theoretical model, and it limits several sources of misspecification risk. We employ the method to estimate the Epstein-Zin preference parameters for an agent representing U.S. consumers and investors in the U.S. equity and T-bill markets from 1952 to 2019. The point estimates of risk aversion and EIS parameters are 16 and 1.7, respectively, close to previously theorized values. Confidence intervals for these estimates indicate a preference for an early risk resolution.

Studies in finance and macroeconomics differ in specifying the agents' recursive preferences and their information. As our estimation approach exploits solely and entirely each of these specifications, the method can be included in a model selection procedure. We leave this extension to future work.

References

Backus, D. K., Routledge, B. R., and Zin, S. E. (2004). Exotic preferences for macroeconomists. *NBER Macroeconomics Annual*, 19:319–390.

Bansal, R., Kiku, D., and Yaron, A. (2010). Long run risks, the macroeconomy, and asset prices. *American Economic Review*, 100(2):542–46.

Bansal, R., Kiku, D., and Yaron, A. (2012). An empirical evaluation of the long-run risks model for asset prices. *Critical Finance Review*, 1:183–221.

Bansal, R., Kiku, D., and Yaron, A. (2016). Risks for the long run: estimation with time aggregation. *Journal of Monetary Economics*, 82:52–69.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481–1509.

Barnett, M., Brock, W., and Hansen, L. P. (2020). Pricing uncertainty induced by climate change. *The Review of Financial Studies*, 33(3):1024–1066.

Borovička, J. and Stachurski, J. (2020). Necessary and sufficient conditions for existence and uniqueness of recursive utilities. *The Journal of Finance*, 75(3):1457–1493.

Brown, A. L. and Kim, H. (2014). Do individuals have preferences used in macro-finance models? An experimental investigation. *Management Science*, 60(4):939–958.

Campanale, C., Castro, R., and Clementi, G. L. (2010). Asset pricing in a production economy with Chew-Dekel preferences. *Review of Economic Dynamics*, 13(2):379–402.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228.

Chen, X., Favilukis, J., and Ludvigson, S. C. (2013). An estimation of economic models with recursive preferences. *Quantitative Economics*, 4(1):39–83.

Chew, S. H. and Epstein, L. G. (1989). The structure of preferences and attitudess towards the timing of the resolution of uncertainty. *International Economic Review*, pages 103–117.

Christensen, T. M. (2017). Nonparametric stochastic discount factor decomposition. *Econometrica*, 85(5):1501–1536.

Christensen, T. M. (2022). Existence and uniqueness of recursive utilities without boundedness. *Journal of Economic Theory*, 200:105413.

Constantinides, G. M. and Ghosh, A. (2011). Asset pricing tests with long-run risks in consumption growth. *The Review of Asset Pricing Studies*, 1(1):96–136.

Dekel, E. (1986). An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. *Journal of Economic Theory*, 40(2):304–318.

Domínguez, M. A. and Lobato, I. N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, 72(5):1601–1615.

Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969.

Epstein, L. G. and Zin, S. E. (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy*, 99(2):263–286.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Gagliardini, P. and Ronchetti, D. (2020). Comparing asset pricing models by the conditional hansen-jagannathan distance. *Journal of Financial Econometrics*, 18(2):333–394.

Gospodinov, N. and Otsu, T. (2012). Local GMM estimation of time series models with conditional moment restrictions. *Journal of Econometrics*, 170(2):476–490.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Hansen, L. P., Heaton, J. C., and Li, N. (2008). Consumption strikes back? Measuring long-run risk. *Journal of Political Economy*, 116(2):260–302.

Hansen, L. P. and Scheinkman, J. A. (2009). Long-term risk: an operator approach. *Econometrica*, 77(1):177–234.

Hansen, L. P. and Scheinkman, J. A. (2012). Recursive utility in a Markov environment with stochastic growth. *Proceedings of the National Academy of Sciences*, 109(30):11967–11972.

Hansen, L. P. and Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, 50(5):1269–1286.

Judd, K. L. (1998). *Numerical methods in economics*. MIT Press.

Kleibergen, F. and Zhan, Z. (2020). Robust inference for consumption-based asset pricing. *The Journal of Finance*, 75(1):507–550.

Koopmans, T. C. (1960). Stationary ordinal utility and impatience. *Econometrica*, 28(2):287–309.

Kreps, D. M. and Porteus, E. L. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46(1):185–200.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance*, 56(3):815–849.

Manresa, E., Peñaranda, F., and Sentana, E. (2017). Empirical evaluation of overspecified asset pricing models. *Working Paper*.

Mehra, R. and Prescott, E. C. (1985). The equity premium: a puzzle. *Journal of Monetary Economics*, 15(2):145–161.

Mumtaz, H. and Theodoridis, K. (2020). Dynamic effects of monetary policy shocks on macroeconomic volatility. *Journal of Monetary Economics*, 114:262–282.

Nagel, S. and Singleton, K. J. (2011). Estimation and evaluation of conditional asset pricing models. *The Journal of Finance*, 66(3):873–909.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Pohl, W., Schmedders, K., and Wilms, O. (2018). Higher order effects in asset pricing models with long-run risks. *The Journal of Finance*, 73(3):1061–1111.

Roll, R. (1977). A critique of the asset pricing theory's tests Part I: on past and potential testability of the theory. *Journal of Financial Economics*, 4(2):129–176.

Routledge, B. R. and Zin, S. E. (2010). Generalized disappointment aversion and asset prices. *The Journal of Finance*, 65(4):1303–1332.

Stock, J. H. and Wright, J. H. (2000). GMM with weak identification. *Econometrica*, 68(5):1055–1096.

Tripathi, G. and Kitamura, Y. (2003). Testing conditional moment restrictions. *The Annals of Statistics*, 31(6):2059–2095.

Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*. CRC Press.

Weber, C. E. (2000). Rule-of-thumb consumption, intertemporal substitution, and risk aversion. *Journal of Business & Economic Statistics*, 18(4):497–502.

Yin, J., Geng, Z., Li, R., and Wang, H. (2010). Nonparametric covariance model.

Statistica Sinica, 20:469.

Yogo, M. (2004). Estimating the elasticity of intertemporal substitution when instruments are weak. *Review of Economics and Statistics*, 86(3):797–810.

Yogo, M. (2006). A consumption-based explanation of expected stock returns. *The Journal of Finance*, 61(2):539–580.

A Proof of Proposition 1

From Assumption 4, we can change the parameterization of the log-consumption growth rate in Equation (5) as a function of $\boldsymbol{Y}_{t+1}$. Similarly, from Assumption (**ii**), we can change the parameterization of the functions $v_{f_\star}$ and $\tilde{v}_f$ in Equations (7) and (8). We have that $\boldsymbol{W}_t(0) = \boldsymbol{Y}_{t,\star}$ and $\tilde{\boldsymbol{W}}_t(0) = \tilde{\mathbf{y}}_*$. Therefore, from Assumptions (**i**) and (**iii**) in Proposition 1, the expectations in the r.h.s.'s of Equations (6) and (9) are such that

$$\mathrm{E}_{[0]}\left[e^{(1-\gamma)g\left(\boldsymbol{W}_{t+1}(0),\tilde{\boldsymbol{W}}_t(0)\right)}v_{[0]}^{\alpha}\left(\tilde{\boldsymbol{W}}_{t+1}(0);\boldsymbol{\theta}\right)\middle|\,\boldsymbol{W}_t,\tilde{\boldsymbol{W}}_t\right] = \mu_{[0]}^{[g]}\left(\tilde{\mathbf{y}}_*;\gamma\right)v_{[0]}^{\alpha}\left(\tilde{\mathbf{y}}_*;\boldsymbol{\theta}\right),$$

and

$$\begin{aligned}\ln\left[\mathrm{E}_{[0]}\left[e^{(1-\gamma)\left(\tilde{v}_{[0]}\left(\tilde{\boldsymbol{W}}_{t+1}(0);\beta,\gamma\right)+g(\boldsymbol{W}_{t+1}(0),\tilde{\boldsymbol{W}}_t(0))\right)}\middle|\,\boldsymbol{W}_t,\tilde{\boldsymbol{W}}_t\right]\right]&\\ = \ln\left[\mathrm{E}_{[0]}\left[e^{(1-\gamma)\tilde{v}_{[0]}\left(\tilde{\boldsymbol{W}}_{t+1}(0);\beta,\gamma\right)}\middle|\,\boldsymbol{W}_t,\tilde{\boldsymbol{W}}_t\right]\right] + \ln\left[\mu_{[0]}^{[g]}\left(\tilde{\mathbf{y}}_*;\gamma\right)\right].&\end{aligned}$$

Therefore, we can write Equation (6) as $v_{[0]}\left(\tilde{\mathbf{y}}_*;\boldsymbol{\theta}\right) = 1-\beta+\beta\mu_{[0]}^{[g]}\left(\tilde{\mathbf{y}}_*;\gamma\right)^{\frac{1}{\alpha}}v_{[0]}\left(\tilde{\mathbf{y}}_*;\boldsymbol{\theta}\right)$ and Equation (9) as $\tilde{v}_{[0]}\left(\tilde{\mathbf{y}}_*;\beta,\gamma\right) = \beta\tilde{v}_{[0]}\left(\tilde{\mathbf{y}}_*;\beta,\gamma\right) + \dfrac{\beta}{1-\gamma}\ln\left[\mu_{[0]}^{[g]}\left(\tilde{\mathbf{y}}_*;\gamma\right)\right]$. The proposition follows from reordering the terms.

B Alternative SDF characterization

The true value of the SDF parameter vector is $\boldsymbol{\theta}_\star = \underset{\boldsymbol{\theta}\in\Theta_{f_\star}}{\operatorname{argmin}}\ \mathcal{Q}_{f_\star^{[M]},f_\star}(\boldsymbol{\theta})$, where

$$\mathcal{Q}_{f^{[M]},f}(\boldsymbol{\theta}) := \mathrm{E}_{f^{[M]}}\left[\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)\boldsymbol{e}_f(\boldsymbol{X}_t;\boldsymbol{\theta})'\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}\boldsymbol{e}_f(\boldsymbol{X}_t;\boldsymbol{\theta})\right], \tag{B.1}$$

for the $\boldsymbol{X}_t$-measurable weighting matrix function $\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)$ estimated by $\hat{\boldsymbol{\Omega}}_t$, and where $\mathrm{E}_{f^{[M]}}[\cdot]$ is the unconditional expectation operator for the marginal probability density function $f^{[M]}$. Under Assumption 5, we have the singleton

$$\{\boldsymbol{\theta}_\star\} := \{\boldsymbol{\theta} : \mathrm{E}_{f_\star}\left[\boldsymbol{h}_{f_\star}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta})|\boldsymbol{X}_t\right] = \mathbf{0}_q\}, \tag{B.2}$$

with the pricing error functional serving as a generalized residual. We can write the criterion function in Equation (B.1) for the true pdf's $f_\star^{[M]}$ and $f_\star$ as

$$\begin{aligned}\mathcal{Q}f_\star^{[M]}, f_\star(\boldsymbol{\theta}) := \mathrm{E}_{f_\star^{[M]}}\Big[\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)\mathrm{E}_{f_\star}\left[\boldsymbol{h}_{f_\star}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta})\big|\boldsymbol{X}_t\right]' \\ \boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}\mathrm{E}_{f_\star}\left[\boldsymbol{h}_{f_\star}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta})\big|\boldsymbol{X}_t\right]\Big].\end{aligned} \tag{B.3}$$

Under Assumption 5, we have $\mathcal{Q}_{f_\star^{[M]}, f_\star}(\boldsymbol{\theta}_\star) = 0$, so that $\nabla_{\boldsymbol{\theta}'}\mathcal{Q}_{f_\star^{[M]}, f_\star}(\boldsymbol{\theta}) = \mathbf{0}_p$ locally in an open neighborhood of $\boldsymbol{\theta}_\star$. This equation represents the local identification condition of the SDF parameter. In general, it may hold without Assumption 5. However, that assumption makes the identification global.

C Regularity assumptions

Assumption 7. *The vector $\boldsymbol{X}_t$ is geometrically strong mixing. The stationary pdf f_Z of the vector $\boldsymbol{Z}_t := [\boldsymbol{X}_t' \ \boldsymbol{X}_{t-1}']'$ is of differentiability class $C^r(\mathbb{R}^{2d^{[X]}})$, for integer $r \geqslant 2$, such that $f_Z > 0$ in the interior of the set $\mathcal{X}^2$. The same conditions are satisfied by $f_\star^{[M]}$. Also, $\int_{\mathcal{X}}\int_{\mathcal{X}}\left[\frac{f_Z(\tilde{\boldsymbol{x}}, \boldsymbol{x})}{f_\star^{[M]}(\tilde{\boldsymbol{x}})f_\star^{[M]}(\boldsymbol{x})}\right]^q f_Z(\tilde{\boldsymbol{x}}, \boldsymbol{x})d\tilde{\boldsymbol{x}}d\boldsymbol{x} < \infty$ for real $q > 1$.*

Assumption 8. *There exists a growing sequence of sets $\mathcal{X}_T \subset \mathcal{X}$ with complement $\mathcal{X}_T^C$, for $T \in \mathbb{N}$, and real constants $c_1, c_2 > 0$ such that $\sup_{\boldsymbol{x}\in\mathcal{X}_T}\mathbb{P}\left[\boldsymbol{X}_{t+1} \in \mathcal{X}_T^C|\boldsymbol{X}_t = \boldsymbol{x}\right] \to 0$ as $T \to \infty$, and $\inf_{\boldsymbol{x},\tilde{\boldsymbol{x}}\in\mathcal{X}_T} f_Z(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \geqslant \frac{c_1}{\log[T]^{c_2}}$ and $\inf_{\boldsymbol{x}\in\mathcal{X}_T} f_\star^{[M]}(\boldsymbol{x}) \geqslant \frac{c_1}{\log[T]^{c_2}}$.*

Assumption 9. *The set $\bar{\mathcal{X}}$ is compact and independent of the value $\boldsymbol{\theta}_\star$ of the SDF parameter vector and the time series sample size T. It belongs to the interior of set $\mathcal{X} \subseteq \mathbb{R}^{d^{[X]}}$, and it is such that $\inf_{\boldsymbol{x}\in\bar{\mathcal{X}}} f_\star^{[M]}(\boldsymbol{x}) > 0$.*

Assumption 10. *The parameter $\boldsymbol{\theta}_\star$ is in the interior of the compact set $\Theta_{f_\star} \subset \mathbb{R}^p$.*

Assumption 11. *We have **(i)** $\mathrm{E}_{f_\star^{[M]}}\left[|m_{f_\star}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star)|^{2\tilde{p}}\right] < \infty$, for real $\tilde{p} > 1$ such that $1/\tilde{p} + 1/q = 1$, where $q > 1$ is defined in Assumption 7; and **(ii)** $\sup_{\substack{\boldsymbol{\theta}\in\Theta_{f_\star} \\ \boldsymbol{x}\in\mathcal{X}}} \mathrm{E}_{f_\star}\left[|m_{f_\star}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta})|^{2+d}\big|\boldsymbol{X}_t = \boldsymbol{x}\right] < \infty$, for real $d > 0$.*

Assumption 12. *The kernel function K is a bounded and Lipschitz function on $\mathcal{X}$*

such that $K(\boldsymbol{x}) = K(-\boldsymbol{x}) \geqslant 0$ *for any* $\boldsymbol{x} \in \mathcal{X}$, $\int_{\mathcal{X}} K(\boldsymbol{u}) d\boldsymbol{u} = 1$, $\boldsymbol{K}(\mathbf{0}_{d^{[X]}}) < \infty$, *and* $\int_{\mathcal{X}} \|\boldsymbol{u}\|^r K(\boldsymbol{u}) d\boldsymbol{u} < \infty$, *where* r *is defined in Assumption 7, and* $\int_{\mathbb{R}^{d^{[X]}}} \boldsymbol{u}^{\boldsymbol{i}} K(\boldsymbol{u}) d\boldsymbol{u} = 0$, *for any multi-index* $\boldsymbol{i} \in \mathbb{N}^{d^{[X]}}$ *such that* $|\boldsymbol{i}| \leqslant r - 1$.

Assumption 13. *The bandwidth* $b_T^{[1]} > 0$ *used for the local constant kernel estimation is such that* ***(i)*** $b_T^{[1]} = o(1)$, ***(ii)*** $\dfrac{\log [T]^2}{T b_T^{[1] 3 d^{[X]}}} = o(1)$, *and* ***(iii)*** $T b_T^{[1] 2r} = o(1)$, *where* r *is defined in Assumption 7.*

Assumption 14. *For any* $\boldsymbol{\theta} \in \Theta_{f_\star}$, *the positive integer number* $N_T(\boldsymbol{\theta})$ *used in Proposition 1 is such that* $\lim_{T \to \infty} N_T(\boldsymbol{\theta}) = \infty$, $\lim_{T \to \infty} N_T(\boldsymbol{\theta}) b_T^{[1] d^{[X]}} = \infty$ *for the bandwidth* $b_T^{[1]}$ *in Assumption 13.*

Assumption 15. *The functions* $\mu_{[i]}^{[g]}$ *in Equation* (12) *are of differentiability class* $C^1\left(\tilde{\mathcal{X}} \times \mathbb{R}_+\right)$, *and the functions* $\sigma_{[i]}^{[g]2} : \mathcal{X} \times \mathbb{R}^{d^{[\boldsymbol{\theta}]}} \to \mathbb{R}_+$ *and* $\tilde{\sigma}_{[i]}^{[g]2} : \mathcal{X} \times (0 : 1) \times \mathbb{R}_+ \to \mathbb{R}_+$ *defined as*

$$\sigma_{[i]}^{[g]2}(\tilde{\boldsymbol{x}}; \boldsymbol{\theta}) := \mathrm{V}_{[i]}\left[e^{(1-\gamma) g\left(\boldsymbol{W}_{t+1}(i), \tilde{\boldsymbol{W}}_t(i)\right)} v_{[i]}\left(\tilde{\boldsymbol{W}}_{t+1}(i); \boldsymbol{\theta}\right)^{\alpha} \middle| \tilde{\boldsymbol{W}}_t(i) = \tilde{\boldsymbol{x}} \right]$$
$$\tilde{\sigma}_{[i]}^{[g]2}(\tilde{\boldsymbol{x}}; \beta, \gamma) := \mathrm{V}_{[i]}\left[e^{(1-\gamma)\left(\tilde{v}_{[i-1]}\left(\tilde{\boldsymbol{W}}_{t+1}(i); \beta, \gamma\right) + g(\boldsymbol{W}_{t+1}(i), \tilde{\boldsymbol{W}}_t(i))\right)} \middle| \tilde{\boldsymbol{W}}_t(i) = \tilde{\boldsymbol{x}} \right],$$

for any non-negative integer i, *are of differentiability class* $C^1\left(\tilde{\mathcal{X}} \times \mathbb{R}^{d^{[\boldsymbol{\theta}]}}\right)$ *and* $C^1\left(\tilde{\mathcal{X}} \times (0 : 1) \times \mathbb{R}_+\right)$, *respectively.*

Assumption 16. *The matrix* $\boldsymbol{\Omega}_T(\boldsymbol{X}_t)$ *converges in probability to the positive-definite matrix* $\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)$ *introduced in Equation* (B.1).

Assumption 17. *The* $(q \times d^{[\boldsymbol{\theta}]})$*-dimensional, finite-valued matrix* $J_{[\boldsymbol{\theta}]}(\boldsymbol{X}_t) := \nabla_{\boldsymbol{\theta}'} \mathrm{E}_{f_\star}\left[\boldsymbol{h}_{f_\star}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta}) \,\middle|\, \boldsymbol{X}_t\right] \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_\star}$ *is full column-rank.*

Assumptions 7-9 are on the probabilistic properties of the vector $\boldsymbol{X}_t$. These assumptions simplify the derivation of the large sample properties of our estimators. Relaxing them would add a technical burden. The first part of Assumption 7 allows the application of Central Limit Theorems for sums. The last part of Assumption 7 restricts the dependence between $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t-1}$ at the boundaries of their support. The assumption on the set $\mathcal{X}_T$ makes the densities bounded away from zero from below on $\mathcal{X}_T$ and $\mathcal{X}_T^2$, respectively, at an inverse logarithmic rate as T increases.

We use them to control for boundary effects in the kernel regression and trim the support of the state variables in Proposition 3. Under Assumption 8, the stationary densities of $\boldsymbol{X}_t$ and $[\boldsymbol{X}_t' \ \boldsymbol{X}_{t-1}']'$ are constrained at the boundary of their supports.[21] Assumptions 10 and 11 are on the true value of the model parameters, and Assumptions 12 and 13 concern the kernel and the bandwidth. Function K is a kernel of order r, the same as the differentiability order of the densities in Assumption 7. Condition (**ii**) in Assumption 13 ensures that the second-order terms in the Fréchet expansions are negligible asymptotically (see the proof of Proposition 3). Condition (**iii**) in the same assumption makes the bias of estimators constructed by averaging kernel regression estimators over the conditioning value asymptotically negligible. Assumptions 14 and 15 concern the initial functions in the continuation value reconstruction procedure of Section 2. Under Assumption 16, the criterion we minimize to estimate the SDF parameter vector converges to a function with a unique global minimum. Assumptions 17 constrain the joint probability distribution of returns and state variables.

D Large sample properties of the estimators

We report the asymptotic statistical properties under the regularity assumptions listed in Appendix C. We present the results under the standard simplifying choice of a diagonal bandwidth matrix with equal bandwidths. As is customary in the literature on local nonparametric regressions, we parameterize the bandwidth matrix $\boldsymbol{H}^{[1]}$ by the scalar bandwidth $b_T^{[1]}$ for the matrix $b_T^{[1]}$, and the set $\mathcal{H}^{[2]}$ by the scalar bandwidth $b_T^{[2]}$, so that $K_{b_T^{[1]}}(z) = b_T^{[1]\ -d^{[X]}} K\left(b_T^{[1]\ -1} z\right)$.

D.1 The reconstruction of the SDF family

As explained in Section 2, we estimate the continuation for the SDF parameter value $\boldsymbol{\theta}$ based on a sequence of auxiliary models approximating more closely the one of interest as i approaches the integer $N_T(\boldsymbol{\theta})$ defined in Expression (17). In doing so, we consider a sequential approximation of the continuation value, taking the kernel transition density $\hat{f}_T$ with the bandwidth matrix $\boldsymbol{H}^{[1]}$ in place of the true transition

[21]Christensen (2022) shows, by imposing thin-tail restrictions, that v and $\tilde{v}$ exhibit contraction for an unbounded state variable space $\mathcal{X}$ as well.

density $f_\star$ relying on the contraction mapping in Equation (6). The pricing error functional $\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta}, \boldsymbol{H}_\star^{[1]}\right)$ is the one we reconstruct. Therefore, in place of the unfeasible set in Equation (B.2), we consider the function $\boldsymbol{\rho}$ of the SDF parameter vector that solves the pricing restriction for the auxiliary model. We identify the corresponding value by the related auxiliary no-arbitrage restriction in Assumption D.1. As explained in Section 2, we estimate the expectation operator $\mathrm{E}_{\hat{f}_T}\left[\cdot\middle|\, \boldsymbol{X}_t; \mathcal{H}_\star^{[2]}\right]$ by a local linear kernel regression. Under Assumption D.1, the true value $\boldsymbol{\theta}_\star$ of the SDF parameter vector, which Equation (B.2) defines, is uniquely identified by the auxiliary model based on the auxiliary parameter value $\hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}_\star)$ obtained through the injective function $\hat{\boldsymbol{\rho}}_T : \Theta_{f_\star} \to \hat{\Theta}_T$ that solves the equation

$$\mathrm{E}_{\hat{f}_T}\left[\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}_\star), \boldsymbol{H}_\star^{[1]}\right) \middle| \boldsymbol{X}_t; \mathcal{H}_\star^{[2]}\right] = \mathbf{0}_q. \tag{D.1}$$

By the chain rule for differentiation and Assumption D.1, we express the matrix $\nabla_{\boldsymbol{\theta}} \mathrm{E}_{\hat{f}_T}\left[\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}), \boldsymbol{H}_\star^{[1]}\right) \middle| \boldsymbol{X}_t; \mathcal{H}_\star^{[2]}\right]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\star}$ as $J_{[\boldsymbol{\theta}]}(\boldsymbol{X}_t)\nabla_{\boldsymbol{\theta}}\hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}_\star)$, where the matrix $J_{[\boldsymbol{\theta}]}(\boldsymbol{X}_t)$ is defined in Assumption 17. Intuitively, the lower the gradient of the function $\hat{\boldsymbol{\rho}}_T$, the harder the inference on $\boldsymbol{\theta}_\star$ based on our continuation value reconstruction method. In the extreme case of a flat function $\hat{\boldsymbol{\rho}}_T$, at least two columns of the matrix $M_{\boldsymbol{\theta}}$ are linearly dependent. By uniform convergence of kernel estimators (see, e.g., Hansen (2008)) and Assumptions 9, 7, 8-13, and 11, $\operatorname*{plim}_{T\to\infty} \nabla_{\boldsymbol{\theta}} \mathrm{E}_{\hat{f}_T}\left[\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}), \boldsymbol{H}_\star^{[1]}\right) \middle| \boldsymbol{X}_t; \mathcal{H}_\star^{[2]}\right]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\star} = J_{[\boldsymbol{\theta}]}(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathcal{X}$, and consequently

$$\operatorname*{plim}_{T\to\infty} \nabla_{\boldsymbol{\theta}}\hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}_\star) = \mathbf{I}_{d^{[\boldsymbol{\theta}]}\times d^{[\boldsymbol{\theta}]}}. \tag{D.2}$$

Let us consider the kernel transition density $\hat{f}_T$, kernel bandwidth $b_T^{[1]} \in \mathbb{R}_+$ and the known functions $v_{[0]}$ and $\tilde{v}_{[0]}$ for any $\boldsymbol{\theta} \in \mathbb{R}^{d^{[\boldsymbol{\theta}]}}$, so that, for any $\boldsymbol{\theta} \in \hat{\Theta}_T$ and $\tilde{\boldsymbol{x}} \in \mathcal{X}$, the function

$$\hat{\varphi}_{[i]}\left(\tilde{\boldsymbol{x}}; \boldsymbol{\theta}, b_T^{[1]}\right) := \sum_{t=1}^{T-1} w_T\left(\tilde{\boldsymbol{x}}, \boldsymbol{W}_t(i); b_T^{[1]}\right) e^{(1-\gamma)g\left(\boldsymbol{W}_{t+1}(i), \tilde{\boldsymbol{W}}_t(i)\right)} v_{[i]}\left(\tilde{\boldsymbol{W}}_{t+1}(i); \boldsymbol{\theta}\right)^{\alpha}$$

is the local constant kernel regression estimator of the function

$$\varphi_{[i]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right) := \mathrm{E}_{[i]}\left[e^{(1-\gamma)g\left(\boldsymbol{W}_{t+1}(i),\tilde{\boldsymbol{W}}_t(i)\right)} v_{[i]}\left(\tilde{\boldsymbol{W}}_{t+1}(i);\boldsymbol{\theta}\right)^{\alpha}\middle|\tilde{\boldsymbol{W}}_t(i)=\tilde{\boldsymbol{x}}\right].$$

For $\psi = 1$, we define its counterpart similarly as

$$\tilde{\varphi}_{[i]}\left(\tilde{\boldsymbol{x}};\beta,\gamma\right) := \mathrm{E}_{[i]}\left[e^{(1-\gamma)\left(\tilde{v}_{[i-1]}\left(\tilde{\boldsymbol{W}}_{t+1}(i);\beta,\gamma\right)+g\left(\boldsymbol{W}_{t+1}(i),\tilde{\boldsymbol{W}}_t(i)\right)\right)}\middle|\tilde{\boldsymbol{W}}_t(i)=\tilde{\boldsymbol{x}}\right].$$

From Assumptions 9-14, we have, for any $\tilde{\boldsymbol{x}} \in \mathcal{X}$, $\boldsymbol{\theta} \in \hat{\Theta}_T$, that

$$\sqrt{Tb_T^{[1]d^{[X]}}}\left(\hat{\varphi}_{[i]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta},b_T^{[1]}\right) - \varphi_{[i-1]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right) + \mathfrak{D}\left[\varphi_{[i]}\right]\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbf{K}\left(\mathbf{0}_{d^{[X]}}\right)\frac{\sigma_{[i]}^{[g]2}(\tilde{\boldsymbol{x}};\boldsymbol{\theta})}{f_\star^{[M]}(\tilde{\boldsymbol{x}})}\right), \quad \text{(D.3)}$$

for the kernel convolution function $\mathbf{K}$ in Proposition (2), the conditional moment function $\sigma_{[0]}^{[g]2}$ in Assumption 15, and the operator $\mathfrak{D}$ defined as $\mathfrak{D}\left[\varphi_{[i]}\right]\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right) := \varphi_{[i-1]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right) - \varphi_{[i]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right)$ and similarly for $v_{[1]}$. For $\psi = 1$, the result is similar, with $\hat{\tilde{\varphi}}_{[0]}\left(\tilde{\boldsymbol{x}};\beta,\gamma,b_T^{[1]}\right)$, $\tilde{\varphi}_{f_\star}\left(\tilde{\boldsymbol{x}};\beta,\gamma\right)$ and $\tilde{\sigma}_{[0]}^{[g]2}\left(\tilde{\boldsymbol{x}};\beta,\gamma\right)$ in Assumption 15 replacing $\hat{\varphi}_{[0]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta},b_T^{[1]}\right)$, $\varphi_{f_\star}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right)$ and $\sigma_{[0]}^{[g]2}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right)$, respectively.

Consider a particular $\boldsymbol{\theta} \in (0:1) \times \mathbb{R}_+^2$. For that value, we have two cases. If the functions $v_{[i]}$ and $\tilde{v}_{[i]}$ feature contraction, that is $\boldsymbol{\theta} \in \Theta_{f_\star}$, we have that $\lim_{T\to\infty}\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\mathfrak{D}\left[\varphi_{[N_T(\boldsymbol{\theta})]}\right](\boldsymbol{x};\boldsymbol{\theta})\right| = 0$, for the integer-valued function N_T implicitly defined in Expression (17). Let us now consider the functions $v_{[i]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right) := 1-\beta+\beta\varphi_{[i]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right)^{\frac{1}{\alpha}}$ and $\hat{v}_{[i]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta},b_T^{[1]}\right) := 1-\beta+\beta\hat{\varphi}_{[i]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta},b_T^{[1]}\right)^{\frac{1}{\alpha}}$. For any $a, b, c \in \mathbb{R}$, the function $a + bx^c : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and independent of T. By an application of the DM to this function and Expression (D.3), we obtain, for any $\tilde{\boldsymbol{x}} \in \mathcal{X}$,

$$\sqrt{Tb_T^{[1]d^{[X]}}}\left(\hat{v}_{[i]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta},b_T^{[1]}\right) - v_{[i-1]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right) + \mathfrak{D}\left[v_{[i]}\right]\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbf{K}\left(\mathbf{0}_{d^{[X]}}\right)\frac{\beta^2}{\alpha^2}\frac{\sigma_{[i]}^{[g]2}(\tilde{\boldsymbol{x}};\boldsymbol{\theta})\varphi_{[i-1]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right)^{2\left(\frac{1}{\alpha}-1\right)}}{f_\star^{[M]}(\tilde{\boldsymbol{x}})}\right). \quad \text{(D.4)}$$

This first case occurs if $\boldsymbol{\theta} \in \Theta_{f_\star}$. The special case of $\Theta_{f_\star} \equiv \emptyset$ corresponds to a preference specification without any acceptable solution. When $\psi = 1$, we take similar steps with $\tilde{v}_{[i]}\left(\tilde{\boldsymbol{x}};\beta,\gamma,b_T^{[1]}\right):=\dfrac{\beta}{1-\gamma}\ln\left[\tilde{\varphi}_{[i]}\left(\tilde{\boldsymbol{x}};\beta,\gamma\right)\right]$ and $\hat{\tilde{v}}_{[1]}\left(\tilde{\boldsymbol{x}};\beta,\gamma,b_T^{[1]}\right):=\dfrac{\beta}{1-\gamma}\ln\left[\hat{\tilde{\varphi}}_{[i]}\left(\tilde{\boldsymbol{x}};\beta,\gamma\right)\right]$. Applying the DM to a scaled logarithmic transformation independent of T and Expression (D.3) adjusted to the case $\psi=1$, we obtain, for any $\tilde{\boldsymbol{x}}\in\mathcal{X}$,

$$\sqrt{Tb_T^{[1]d[X]}}\left(\hat{\tilde{v}}_{[i]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta},b_T^{[1]}\right)-\tilde{v}_{[i-1]}\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right)+\mathfrak{D}\left[\tilde{v}_{[i]}\right]\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right)\right)$$

$$\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\frac{\beta^2}{(1-\gamma)^2}\mathbf{K}\left(\mathbf{0}_{d[X]}\right)\frac{\tilde{\sigma}_{[i]}^{[g]2}(\tilde{\boldsymbol{x}};\boldsymbol{\theta})}{f_\star^{[M]}(\tilde{\boldsymbol{x}})}\frac{1}{\varphi_{[i-1]}^2\left(\tilde{\boldsymbol{x}};\boldsymbol{\theta}\right)}\right). \quad \text{(D.5)}$$

Let us now consider the second case with functions $v_{[i]}$ and $\tilde{v}_{[i]}$ not featuring contraction. This second case realizes whatever value $\boldsymbol{\theta}$ is outside the parameter space $\Theta_{f_\star}$. In this case, we have that $\lim_{T\to\infty}\mathfrak{D}\left[\varphi_{[N_T(\boldsymbol{\theta})]}\right]\left(\boldsymbol{X}_t;\boldsymbol{\theta}\right)$ is not null everywhere. If $\mathfrak{D}\left[\varphi_{[i]}\right]\left(\boldsymbol{X}_t;\boldsymbol{\theta}\right)$ converges to a positive (negative, or null) number, the limit is $+\infty$ ($-\infty$, or 0, respectively). All the values $\boldsymbol{\theta}$ leading to this second case lie in the complement of the parameter space $\Theta_{f_\star}$.

D.2 Proof of Proposition 2

The conditional moment condition in Equation (6) valued at the true transition density $f_\star$ is $\mathcal{A}_{\boldsymbol{\theta},f_\star}\left[v_{f_\star}\left(\cdot;\boldsymbol{\theta}\right)\right]\left(\boldsymbol{X}_t\right)=0$ for any $\boldsymbol{\theta}\in\Theta_{f_\star}$. This condition is the asymptotic limit condition for $\mathcal{A}_{\boldsymbol{\theta},\hat{f}_T}\left[v_{\hat{f}_T}\left(\cdot;\boldsymbol{\theta}\right)\right]\left(\boldsymbol{X}_t\right)=0$ for any $\boldsymbol{\theta}\in\hat{\Theta}_T$. Under the identification condition in Assumption 3, if we replace the function $v_{f_\star}\left(\cdot;\boldsymbol{\theta}\right)$ with any other function, the equation is not null for all the values $\boldsymbol{\theta}\in\Theta_{f_\star}$. The set $\Theta_{f_\star}$ is compact by Assumption 10, and the set $\hat{\Theta}_T$ is compact by its same construction. The function $\mathcal{A}_{\boldsymbol{\theta},f_\star}\left[v_{f_\star}\left(\cdot;\boldsymbol{\theta}\right)\right]\left(\boldsymbol{X}_t\right)$ is a continuous function of $\boldsymbol{\theta}$. Finally, the function $\mathcal{A}_{\boldsymbol{\theta},\hat{f}_T}\left[v_{\hat{f}_T}\left(\cdot;\boldsymbol{\theta}\right)\right]\left(\boldsymbol{X}_t\right)$ converges to the function $\mathcal{A}_{\boldsymbol{\theta},f_\star}\left[v_{f_\star}\left(\cdot;\boldsymbol{\theta}\right)\right]\left(\boldsymbol{X}_t\right)$ uniformly in $\boldsymbol{\theta}\in\hat{\Theta}_T$. This convergence derives from the uniform convergence of kernel estimators (see, e.g., Hansen (2008)) and Assumptions 9, 7, 8-13, and 11. Because of these properties, we have that, for any $\mathbf{x}\in\mathcal{X}$, $\sup_{\boldsymbol{\theta}\in\hat{\Theta}_T}\left\|\mathrm{E}_{\hat{f}_T}\left[\varphi\left(\boldsymbol{X}_{t+1}\right)\middle|\boldsymbol{X}_t=\mathbf{x};\boldsymbol{H}^{[2]}\right]-\mathrm{E}_{f_\star}\left[\varphi\left(\boldsymbol{X}_{t+1}\right)\middle|\boldsymbol{X}_t=\mathbf{x}\right]\right\|=o_p(1)$, for any integrable function $\varphi\left(\boldsymbol{X}_{t+1}\right):\mathcal{X}\to\mathbb{R}$. Then, the function $\mathcal{A}_{\boldsymbol{\theta},\hat{f}_T}\left[v_{\hat{f}_T}\left(\cdot;\boldsymbol{\theta}\right)\right]\left(\boldsymbol{X}_t\right)$ converges to the function $\mathcal{A}_{\boldsymbol{\theta},f_\star}\left[v_{f_\star}\left(\cdot;\boldsymbol{\theta}\right)\right]\left(\boldsymbol{X}_t\right)$ uniformly in $\boldsymbol{\theta}\in\hat{\Theta}_T$. Then, the consistency of the SDF parameter space estimator $\hat{\Theta}_T$ follows from Theorem 2.1 in Newey and McFadden (1994).

We now move to the large sample properties of the estimator of the SDF family, considering the functions $\varphi_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}) := e^{-\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} v_f\left(\boldsymbol{X}_{t+1}; \boldsymbol{\theta}\right)^{\alpha-1}$ and $\tilde{\varphi}_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \beta, \gamma) := e^{-\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t) + (1-\gamma)\tilde{v}_f(\boldsymbol{X}_{t+1}; \beta, \gamma)}$. The SDF in System (10) is

$$m_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}) = \begin{cases} \beta \dfrac{\varphi_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta})}{\mathrm{E}_f \left[\varphi_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}) v_f\left(\boldsymbol{X}_{t+1}; \boldsymbol{\theta}\right) e^{g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} \middle| \boldsymbol{X}_t \right]^{1-\frac{1}{\alpha}}}, \\ \hfill \text{if } \psi \neq 1, \\ \beta \dfrac{\tilde{\varphi}_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \beta, \gamma)}{\mathrm{E}_f \left[\tilde{\varphi}_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \beta, \gamma) e^{g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} \middle| \boldsymbol{X}_t \right]}, \text{ if } \psi = 1. \end{cases}$$

From System (10) and the SDF normalization $\mathrm{E}_{f_\star}\left[m_{f_\star}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) \middle| \boldsymbol{X}_t \right] = 1$, the function $\beta \mathrm{E}_{f_\star}\left[e^{(1-\gamma_\star) g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} v_{f_\star}\left(\boldsymbol{X}_{t+1}; \boldsymbol{\theta}\right)^{\alpha_\star} \middle| \boldsymbol{X}_t \right]^{\frac{1}{\alpha_\star}-1}$ coincides with the function $\mathrm{E}_{f_\star}\left[e^{-\gamma g(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t)} v_{f_\star}\left(\boldsymbol{X}_{t+1}; \boldsymbol{\theta}_\star\right)^{\alpha_\star - 1} \middle| \boldsymbol{X}_t \right]^{-1}$, with $\alpha_\star := (1-\gamma_\star)/(1-1/\psi_\star)$. So, $\mathrm{E}_{\hat{f}_T}\left[m_{\hat{f}_T}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star, b_T^{[1]} \right) \middle| \boldsymbol{X}_t \right] = \tilde{\varepsilon}_{\hat{f}_T}(\boldsymbol{X}_t; \beta_\star, \gamma_\star, b_T^{[1]}) = 1 + \mathcal{O}_p\left(1/\sqrt{T b_T^{[1]}} \right)$. Then, the SDF in Equations (18) and (19) is

$$\begin{aligned} & m_{\hat{f}_T}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star, b_T^{[1]}) \\ & = \begin{cases} \hat{\varphi}_{[N_T(\boldsymbol{\theta})]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) / \, \mathrm{E}_{\hat{f}_T}\left[\hat{\varphi}_{[N_T(\boldsymbol{\theta}_\star)-1]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) \middle| \boldsymbol{X}_t \right], \\ \hfill \text{if } \psi_\star \neq 1, \\ \hat{\tilde{\varphi}}_{[N_T(\boldsymbol{\theta}_\star)]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \beta_\star, \gamma_\star) / \, \mathrm{E}_{\hat{f}_T}\left[\hat{\tilde{\varphi}}_{[N_T(\boldsymbol{\theta}_\star)-1]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \beta_\star, \gamma_\star) \middle| \boldsymbol{X}_t \right], \\ \hfill \text{if } \psi_\star = 1. \end{cases} \end{aligned}$$

with a local constant kernel regression estimator at the denominator. As explained in Section 2, our iterative reconstruction of the functions $v_f\left(\cdot; \boldsymbol{\theta}\right)$ and $\tilde{v}_f\left(\cdot; \beta, \gamma\right)$ start with the functions $v_{[0]}\left(\cdot; \boldsymbol{\theta}\right)$ and $\tilde{v}_{[0]}\left(\cdot; \beta, \gamma\right)$. Together with the log-consumption growth rate $g\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t\right)$, they determine the design of the nonparametric regression. We indeed interpret our parameter inference conditional on the state variables sample. From Assumptions 9-13, we have, for any $\boldsymbol{x} \in \mathcal{X}$, $\boldsymbol{\theta} \in \hat{\Theta}_T$, that

$$\sqrt{Tb_T^{[1]d^{[X]}}} \Bigg(\mathrm{E}_{\hat{f}_T} \left[\hat{\varphi}_{[N_T(\boldsymbol{\theta}_\star)-1]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) \middle| \boldsymbol{X}_t = \boldsymbol{x} \right] - \mathrm{E}_{f_\star} \left[\varphi_{[N_T(\boldsymbol{\theta}_\star)-1]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) \middle| \boldsymbol{X}_t = \boldsymbol{x} \right] \Bigg) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{\mathbf{K}\left(\mathbf{0}_{d^{[X]}}\right) \mathrm{V}_{f_\star} \left[\varphi_{[N_T(\boldsymbol{\theta}_\star)-1]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) \middle| \boldsymbol{X}_t = \boldsymbol{x} \right]}{f_\star^{[M]}(\boldsymbol{x})} \right).$$

Since the function $\zeta_1 : \mathbb{R} \to \mathbb{R}$ defined as $\zeta_1(x) \coloneqq 1/x$ is independent of T, and continuously differentiable with $\zeta_1' = -1/x^2 \neq 0$, we get from the DM that

$$\sqrt{Tb_T^{[1]d^{[X]}}} \left(\mathrm{E}^{-1}_{\hat{f}_T} \left[\hat{\varphi}_{[N_T(\boldsymbol{\theta}_\star)-1]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) \middle| \boldsymbol{X}_t = \boldsymbol{x} \right] - \mathrm{E}^{-1}_{f_\star} \left[\varphi_{[N_T(\boldsymbol{\theta}_\star)-1]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) \middle| \boldsymbol{X}_t = \boldsymbol{x} \right] \right) \xrightarrow{\mathcal{D}} Z,$$

for the random variable

$$Z \sim \mathcal{N} \left(0, \frac{\mathbf{K}\left(\mathbf{0}_{d^{[X]}}\right) \mathrm{V}_{f_\star} \left[\varphi_{[N_T(\boldsymbol{\theta}_\star)-1]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}) \middle| \boldsymbol{X}_t = \boldsymbol{x} \right]}{\mathrm{E}^2_{f_\star} \left[\varphi_{[N_T(\boldsymbol{\theta}_\star)-1]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) \middle| \boldsymbol{X}_t = \boldsymbol{x} \right] f_\star^{[M]}(\boldsymbol{x})} \right).$$

Then, we can approximate

$$m_{\hat{f}_T}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star, b_T^{[1]}) \approx \left(1/\sqrt{Tb_T^{[1]d^{[X]}}} \right) Z \hat{\varphi}_{[N_T(\boldsymbol{\theta}_\star)]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) + \hat{\varphi}_{[N_T(\boldsymbol{\theta}_\star)]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) \, \mathrm{E}^{-1}_{f_\star} \left[\varphi_{[N_T(\boldsymbol{\theta}_\star)-1]}(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star) \middle| \boldsymbol{X}_t \right].$$

We take similar steps for the case $\psi = 1$, with the function $\tilde{\varphi}_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \beta_\star, \gamma_\star)$ in place of $\varphi_f(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t; \boldsymbol{\theta}_\star)$. Proposition 2 follows.

D.3 Proof of Proposition 3

We prove the consistency of the estimator $\hat{\boldsymbol{\theta}}_T$ by checking the Assumptions **(i)-(iv)** of Theorem 2.1 in Newey and McFadden (1994). (**i**) Let us consider the limit criterion $\mathcal{Q}_{f_\star^{[M]}, f_\star}(\boldsymbol{\theta})$ in Equation (B.3) for $\boldsymbol{\theta} \in \Theta_{f_\star}$, that is the asymptotic limit of the criterion $\mathcal{Q}_{\hat{f}_T}$ minimized by $\hat{\boldsymbol{\theta}}_T$ (see Equation (21)). The limit criterion $\mathcal{Q}_{f_\star^{[M]}, f_\star}(\boldsymbol{\theta})$ is uniquely minimized at $\boldsymbol{\theta}_\star$ by the identification condition in Assumption 5 and since $\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)$ is positive-definite (Assumption 16). (**ii**) The set $\Theta_{f_\star}$ is compact by Assumption 10. (**iii**) For any $\mathbf{x} \in \mathcal{X}$, the mapping

$\boldsymbol{\theta} \to \mathrm{E}_{f_\star}\left[\boldsymbol{h}_{f_\star}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta}\right) \middle| \boldsymbol{X}_t = \mathbf{x}\right]$ is continuous. Therefore, the criterion $\mathcal{Q}_{f_\star^{[M]}, f_\star}(\boldsymbol{\theta})$, which is a quadratic form of it, is also continuous. (**iv**) Let us verify that $\mathcal{Q}_{\hat{f}_T}(\boldsymbol{\theta})$ converges to $\mathcal{Q}_{f_\star^{[M]}, f_\star}(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta} \in \hat{\Theta}_T$, and, for this purpose, consider the vector $\mathbb{1}_{\mathcal{X}}(\boldsymbol{X}_{t+1}) \boldsymbol{h}_{f_\star}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta}\right)$. By uniform convergence of kernel estimators (see, e.g., Hansen (2008)) and Assumptions 9, 7, 8-13, and 11,

$$\sup_{\boldsymbol{\theta} \in \hat{\Theta}_T} \left\| \mathrm{E}_{\hat{f}_T}\left[\mathbb{1}_{\mathcal{X}_T}(\boldsymbol{X}_{t+1}) \boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}), \boldsymbol{H}_\star^{[1]}\right) \middle| \boldsymbol{X}_t; \mathcal{H}_\star^{[2]}\right] - \mathrm{E}_{f_\star}\left[\mathbb{1}_{\mathcal{X}}(\boldsymbol{X}_{t+1}) \boldsymbol{h}_{f_\star}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \boldsymbol{\theta}\right) \middle| \boldsymbol{X}_t = \mathbf{x}\right] \right\| = o_p(1) \quad \text{(D.6)}$$

for any $\mathbf{x} \in \mathcal{X}$. where $\mathcal{X}_T$ is in Assumption 8. Then, from Equation (D.6), and Assumption 16, $\mathcal{Q}_{\hat{f}_T}(\boldsymbol{\theta})$ converges to $\mathcal{Q}_{f_\star^{[M]}, f_\star}(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta} \in \hat{\Theta}_T$ for any $\mathbf{x} \in \mathcal{X}$.

We now prove the asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}_T$. The first-order condition for estimator $\hat{\boldsymbol{\theta}}_T$ in Equation (21) is

$$\begin{aligned}
&\sum_{t=1}^{T-1} \mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t) \nabla_{\boldsymbol{\theta}'} \boldsymbol{e}_{\hat{f}_T}\left(\boldsymbol{X}_t; \hat{\boldsymbol{\theta}}_T, \boldsymbol{H}_\star^{[1]}, \mathcal{H}_\star^{[2]}\right)' \hat{\boldsymbol{\Omega}}_T(\boldsymbol{X}_t)^{-1} \boldsymbol{e}_{\hat{f}_T}\left(\boldsymbol{X}_t; \hat{\boldsymbol{\theta}}_T, \boldsymbol{H}_\star^{[1]}, \mathcal{H}_\star^{[2]}\right) \\
&= \sum_{t=1}^{T-1} \mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t) \mathrm{E}_{\hat{f}_T}\left[\nabla_{\boldsymbol{\theta}'} \boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right), \boldsymbol{H}_\star^{[1]}\right) \middle| \boldsymbol{X}_t; \mathcal{H}_\star^{[2]}\right]' \\
&\qquad \hat{\boldsymbol{\Omega}}_T(\boldsymbol{X}_t)^{-1} \mathrm{E}_{\hat{f}_T}\left[\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{R}_{t+1}; \hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right), \boldsymbol{H}_\star^{[1]}\right) \middle| \boldsymbol{X}_t; \mathcal{H}_\star^{[2]}\right] = \mathbf{0}_p.
\end{aligned}$$

By the mean-value theorem, there exists a vector $\hat{\boldsymbol{\rho}}_T\left(\tilde{\boldsymbol{\theta}}\right)$, whose components are singularly between those of $\hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right)$ and $\hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}_\star)$, such that

$$\sum_{t=1}^{T-1}\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)\mathrm{E}_{\hat{f}_T}\Big[\nabla_{\boldsymbol{\theta}'}\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right),\boldsymbol{H}_\star^{[1]}\right)\Big|\boldsymbol{X}_t;\mathcal{H}_\star^{[2]}\Big]'$$
$$\hat{\boldsymbol{\Omega}}_T(\boldsymbol{X}_t)^{-1}\mathrm{E}_{\hat{f}_T}\Big[\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}_\star),\boldsymbol{H}_\star^{[1]}\right)\Big|\boldsymbol{X}_t;\mathcal{H}_\star^{[2]}\Big]$$
$$+\left(\sum_{t=1}^{T-1}\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)\mathrm{E}_{\hat{f}_T}\Big[\nabla_{\boldsymbol{\theta}'}\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right),\boldsymbol{H}_\star^{[1]}\right)\Big|\boldsymbol{X}_t;\mathcal{H}_\star^{[2]}\Big]'\right.$$
$$\left.\hat{\boldsymbol{\Omega}}_T(\boldsymbol{X}_t)^{-1}\mathrm{E}_{\hat{f}_T}\Big[\nabla_{\boldsymbol{\theta}'}\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\hat{\boldsymbol{\rho}}_T\left(\tilde{\boldsymbol{\theta}}\right),\boldsymbol{H}_\star^{[1]}\right)\Big|\boldsymbol{X}_t;\mathcal{H}_\star^{[2]}\Big]\right)$$
$$\left(\hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right)-\hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}_\star)\right)=\mathbf{0}_p.$$

By multiplying the two sides of the last equation by $\sqrt{T-1}$ and rearranging its terms, we get $\sqrt{T-1}\left(\hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right)-\hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}_\star)\right)=-A_T^{-1}B_T$, for the variables

$$A_T:=\frac{1}{T-1}\sum_{t=1}^{T-1}\mathrm{E}_{\hat{f}_T}\Big[\nabla_{\boldsymbol{\theta}'}\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right),\boldsymbol{H}_\star^{[1]}\right)\Big|\boldsymbol{X}_t;\mathcal{H}_\star^{[2]}\Big]'$$
$$\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)\hat{\boldsymbol{\Omega}}_T(\boldsymbol{X}_t)^{-1}\mathrm{E}_{\hat{f}_T}\Big[\nabla_{\boldsymbol{\theta}'}\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\hat{\boldsymbol{\rho}}_T\left(\tilde{\boldsymbol{\theta}}\right),\boldsymbol{H}_\star^{[1]}\right)\Big|\boldsymbol{X}_t;\mathcal{H}_\star^{[2]}\Big],$$

and

$$B_T:=\frac{1}{T-1}\sum_{t=1}^{T-1}\mathrm{E}_{\hat{f}_T}\Big[\nabla_{\boldsymbol{\theta}'}\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right),\boldsymbol{H}_\star^{[1]}\right)\Big|\boldsymbol{X}_t;\mathcal{H}_\star^{[2]}\Big]'$$
$$\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)\hat{\boldsymbol{\Omega}}_T(\boldsymbol{X}_t)^{-1}\mathrm{E}_{\hat{f}_T}\Big[\boldsymbol{h}_{\hat{f}_T}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\hat{\boldsymbol{\rho}}_T(\boldsymbol{\theta}_\star),\boldsymbol{H}_\star^{[1]}\right)\Big|\boldsymbol{X}_t;\mathcal{H}_\star^{[2]}\Big].$$

From the consistency of the kernel regression estimators and Equation (D.2), we have

$$A_T=\mathrm{E}_{f_\star^{[M]}}\left[\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)J_{[\boldsymbol{\theta}]}(\boldsymbol{X}_t)'\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}J_{[\boldsymbol{\theta}]}(\boldsymbol{X}_t)\right]+o_p(1)$$
$$B_T=\frac{1}{\sqrt{T-1}}\sum_{t=1}^{T-1}\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)J_{[\boldsymbol{\theta}]}(\boldsymbol{X}_t)'\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}$$
$$\cdot\,\mathrm{E}_{f_\star}\left[\boldsymbol{h}_{f_\star}(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\boldsymbol{\theta}_\star)\,\middle|\,\boldsymbol{X}_t\right]+o_p(1).$$

Then,

$$\sqrt{T-1}\left(\hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right)-\hat{\boldsymbol{\rho}}_T\left(\boldsymbol{\theta}_\star\right)\right)$$
$$=o_p(1)-\mathrm{E}_{f_\star^{[M]}}\left[\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)'\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)\right]^{-1}$$
$$\cdot\frac{1}{\sqrt{T-1}}\sum_{t=1}^{T-1}\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)'\boldsymbol{\Omega}_\star\left(\boldsymbol{X}_t\right)^{-1}\mathrm{E}_{f_\star}\left[\boldsymbol{h}_{f_\star}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\boldsymbol{\theta}_\star\right)\big|\boldsymbol{X}_t\right].$$

So, $\sqrt{T-1}\left(\hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right)-\hat{\boldsymbol{\rho}}_T\left(\boldsymbol{\theta}_\star\right)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\mathrm{Avar}_{f_\star^{[M]}}\left[\hat{\boldsymbol{\theta}}_T\right]\right)$ from the first part of Assumption 7 and the Central Limit Theorem for stationary and α-mixing variables with the asymptotic variance-covariance matrix

$$\mathrm{Avar}_{f_\star^{[M]}}\left[\hat{\boldsymbol{\theta}}_T\right]:=\mathrm{E}_{f_\star^{[M]}}\left[\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)'\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)\right]^{-1}$$
$$\mathrm{E}_{f_\star^{[M]}}\left[\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)'\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}\,\mathrm{V}_{f_\star}\left[\boldsymbol{h}_{f_\star}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\boldsymbol{\theta}_\star\right)\right|\boldsymbol{X}_t\right]$$
$$\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)\Big]\mathrm{E}_{f_\star^{[M]}}\left[\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)'\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)\right]^{-1}.$$

Asymptotically, there is no bias as $Tb_T^{[1]2r}=o(1)$ in Assumption 13. The asymptotic variance of the estimator $\hat{\boldsymbol{\theta}}_T$ is minimal for the weighting matrix $\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)=\mathrm{V}_{f_\star}\left[\boldsymbol{h}_{f_\star}\left(\boldsymbol{X}_{t+1},\boldsymbol{X}_t,\boldsymbol{R}_{t+1};\boldsymbol{\theta}_\star\right)\right|\boldsymbol{X}_t]$. In that case, the asymptotic variance is equal to $\mathrm{E}_{f_\star^{[M]}}\left[\mathbb{1}_{\bar{\mathcal{X}}}(\boldsymbol{X}_t)J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)'\boldsymbol{\Omega}_\star(\boldsymbol{X}_t)^{-1}J_{[\boldsymbol{\theta}]}\left(\boldsymbol{X}_t\right)\right]^{-1}$. The asymptotic behavior of $\hat{\boldsymbol{\rho}}_T\left(\hat{\boldsymbol{\theta}}_T\right)$, Equation (D.2), and the implications for stochastic convergence modes prove Proposition 3.