Estimating Dynamic Equilibrium Models using Mixed Frequency Macro and Financial Data*

Bent Jesper Christensen[(a,b)], Olaf Posch[(c,b)†], and Michel van der Wel[(b,d)]
[(a)]Aarhus University, [(b)]CREATES, [(c)] Universität Hamburg,
[(d)]Erasmus University Rotterdam

April 18, 2016

Abstract

We provide a framework for inference in dynamic equilibrium models including financial market data at daily frequency, along with macro series at standard lower frequency. Our formulation of the macro-finance model in continuous time conveniently accounts for the difference in observation frequency. We suggest the use of martingale estimating functions (MEF) to infer the structural parameters of the model directly through a nonlinear scheme. This method is compared to regression-based methods and the generalized method of moments (GMM). We illustrate our approaches by estimating various versions of the AK-Vasicek model with mean-reverting interest rates. We provide asymptotic theory and Monte Carlo evidence on the small sample behavior of the estimators and report empirical estimates using 30 years of U.S. macro and financial data.

JEL classification: C13; E32; O40

Keywords: Structural estimation; AK-Vasicek model; Martingale estimating function

*We are grateful to two anonymous referees, the editors, Anton Braun, Fabio Canova, Eric Ghysels, Andrew Harvey, Roderick McCrorie, Hashem Pesaran, Juan Rubio-Ramírez, Frank Schorfheide, Frank Smets, seminar participants in Mannheim, Groningen, Vienna, St. Andrews, and conference participants in Chicago (ASSA, SoFiE), Santa Fe (WFA), Oslo (ESEM), Florence (Econometric Modelling of Macro-Financial Linkages), Göttingen (VfS), Atlanta (Methods and Applications for DSGE Models), Vancouver (CEF), Nicosia (EC2), Aarhus (Recent Developments on the Links between Financial Markets and the Real Economy), Berlin (Humboldt-Copenhagen), Tilburg (NESG), and London (IAAE) for comments. The authors appreciate financial support from the Netherlands Organisation for Scientific Research (NWO) for a Veni grant, from the Danish Social Science Research Council, and from the Center for Research in Econometric Analysis of TimE Series (CREATES), which is funded by the Danish National Research Foundation.

†Corresponding author: Olaf Posch (Email address: olaf.posch@uni-hamburg.de, Address: Universität Hamburg, Department of Economics, Von-Melle-Park 5, 20146 Hamburg, Germany).

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models have become the workhorse in macroeconomics, capturing aggregate dynamics over the business cycle. They are frequently used in academic research and to assess various policy interventions. Given the importance and relevance of these models in theoretical and applied research, surprisingly little work has been focused on reconciling business cycle facts with asset pricing implications (see the results in Jermann, 1998; Tallarini, 2000; Rudebusch and Swanson, 2008). The 2007 economic and credit crises intensified the desire to link macro and finance.[1] Conceptually, financial models should benefit from specifying the stochastic discount factor consistently with macro-dynamics, whereas macroeconomic models could benefit from rich financial data.[2] Essential features of this strand of literature are the macro-finance interaction, the mixed frequency of macroeconomic and financial data, and latent variables. Yet, there is no clear consensus on how macro and financial data should be linked consistently, and how data are used efficiently in the estimation of macro-finance models.

Our aim is to develop a paradigm in which macroeconomics, finance, and econometrics are coherently linked. This paper provides a framework for estimation of dynamic equilibrium models with both macro and financial variables, taking account of mixed frequencies and latent variables. We believe that a structural estimation approach can shed light on the channels through which financial markets and the real economy interact. This is important for the design of monetary and fiscal policies and to evaluate policy measures. Specifically, our contribution is threefold. First, we propose using DSGE models in continuous time to facilitate incorporation of financial market variables in a structural manner. Second, we propose and develop martingale estimating functions (MEF) for such macro-finance models. Our continuous-time formulation and the MEF approach naturally accommodate variables arriving at different frequencies by a model-based time-aggregation and thus facilitates statistical inference in the presence of mixed-frequency data. Third, we extend the baseline MEF approach to further consider the case of mixed-frequency data estimation and additionally develop techniques for estimation of dynamic equilibrium models in case of latent

[1] Recent developments illustrate that such a unified framework is promising and intriguing: Gertler and Karadi (2011) present a DSGE model with financial frictions in which intermediaries face balance sheet constraints. Hence, unconventional monetary policy, expanding central bank credit intermediation, may serve as a complement to financial intermediation. Brunnermeier and Sannikov (2014) develop a macroeconomic model with a financial sector and endogenous leverage which leads to crisis episodes, showing a mechanism for how small shocks can have potentially large effects on the real economy.

[2] There are developments incorporating macro factors in financial models of the term structure (Ang and Piazzesi, 2003; Dewachter and Lyrio, 2006; Diebold, Rudebusch, and Aruoba, 2006; Hördahl, Tristani, and Vestin, 2006; Rudebusch and Wu, 2008) and incorporating financial factors in the estimation of macro models (Ang, Piazzesi, and Wei, 2006; van Binsbergen, Fernández-Villaverde, Koijen, and Rubio-Ramírez, 2012).

variables by a simulation-based MEF (SMEF).

We make the link between macro and financial markets explicit by showing how financial market data (say, interest rates or return data) can be used to identify the structural parameters characterizing preferences and technology. To make the least stringent timing assumption we cast our DSGE model in continuous time, solve for the general equilibrium of the real economy and asset prices, and then develop three alternative estimation procedures. We consider off-the-shelf regression-based methods (combined with minimum distance methods to identify structural parameters), the generalized method of moments (GMM), and martingale estimating functions (MEF). Our continuous-time formulation is useful in three respects: (i) to place structure on the residuals in the regression-based methods, (ii) to obtain the general equilibrium dynamics in terms of data and parameters for the GMM and MEF approaches, and (iii) to account for the dependence among variables during the observation interval.

Our analysis is motivated by the fact that financial market data typically are available at higher frequency and of better quality than aggregate macro data (e.g., they are not subject to revision). In a unified framework, financial variables provide an additional source of evidence on the state of the economy, beyond macro series. So far only a few researchers have made use of this property in DSGE models. One apparent challenge is that discrete-time models are not time invariant (see Marcellino, 1999; Foroni and Marcellino, 2014). Put differently, the parameter estimates can only be properly interpreted in context of the particular way in which we solve our models and given the particular rate at which we sample the underlying process.[3] But any fixed period length is arbitrary. Moreover, the frequency at which economic agents make their decisions may not necessarily coincide with the observation frequency of either macro or financial variables. Formulating structural models in continuous time offers a way forward. Our approach yields explicit functional forms for the relations among observables without taking a stand on the frequency at which economic agents make their decisions. Having at hand these functional forms, the availability of financial data at higher frequency (say, daily) than consumption and production (monthly or quarterly) then allows precise approximation of integrals by summation over days. Because the structural parameters enter into the coefficients on these terms, financial data improve identification. Our results indicate that our approach may help resolve the lack of identification of some parameters in the equivalent (log-linearized) discrete-time model (cf. Canova and Sala, 2009). We also show how the MEF approach can be extended to cope with (more

[3]The temporal aggregation literature started with Amemiya and Wu (1972); Geweke (1978). Others noted how the time-invariance problem relates to the behavior of agents (Christiano and Eichenbaum, 1987). Recent contributions include Kim (2010) and Giannone, Monti, and Reichlin (2014).

general) mixed-frequency data and latent variables, including regime-switching models and stochastic volatility.

We depart from the traditional discrete-time formulation of DSGE models and their estimation for three related technical reasons.[4] First, there is no need to perform numerical integration to compute expectations, since the Hamilton-Jacobi-Bellman (HJB) equation is non-stochastic, thus simplifying computation of the first-order principles. Second, some solutions allow for an analytical likelihood function, simplifying the inference on structural parameters even in the presence of nonlinearities and non-normalities (see Posch, 2009). Third, many seminal models in finance are stated in continuous time (such as the equilibrium models of Vasicek, 1977; Cox, Ingersoll, and Ross, 1985a), which is particulary useful in the development of a unified framework in macroeconomics and finance.

Our work builds on a tradition in macroeconomics estimating continuous-time models formulated as systems of (linear) stochastic differential equations.[5] The traditional approach is to solve the system, leading to a coefficient matrix that is a function of the exponential of a matrix depending on the structural model parameters (cf. Phillips, 1972). As illustrated in McCrorie (2009), this complicates identification due to the aliasing phenomenon: The distinct stochastic processes may look identical when sampled at discrete intervals (see Hansen and Scheinkman, 1995, p. 769).[6] In this paper, we adopt an alternative approach of integrating the logarithmic (nonlinear) system to get an 'exact' discrete-time analog. The resulting system is in logarithmic growth rates rather than levels. It involves a coefficient matrix linear in a set of known functions of the structural parameters, and does not involve any matrix exponential. An analysis of whether our approach alleviates the aliasing problem in the linear model is interesting, but beyond the scope of this paper.

Our martingale estimating functions (MEF) approach benefits from the continuous-time structure of the dynamic equilibrium model when computing conditional expectations through deterministic Taylor expansions or using the mixed-frequency properties of the data, but may also be applied to discrete-time models. In either case, the models produce mar-

[4]An non-exhaustive list of references on the estimation of discrete-time DSGE models is Ruge-Murcia (2007); Fernández-Villaverde and Rubio-Ramírez (2007), and An and Schorfheide (2007). While the first two references show how to use standard econometric methods and the particle filter (instead of the linear Kalman filter) for estimation, the latter reviews Bayesian methods. In an accompanying web appendix we discuss in Section B.4 how an Euler approximation could be used to apply this toolbox to continuous-time DSGE models. We do *not* follow this route because the continuous-time formulation naturally accounts for the different observation frequencies of macro and financial market data, a benefit which in this case would be lost.

[5]Seminal papers are Bergstrom (1966); Sims (1971); Phillips (1972, 1991), along with the contributions on rational expectations models by Hansen and Sargent (1991); Hansen and Scheinkman (1995).

[6]Hence, one may argue that the use of continuous-time models is not a panacea. What is gained in terms of time invariance may come at the cost of needing to solve a more severe econometric identification problem.

tingale increments, and the martingale estimating functions are defined as weighted sums of these. The optimal weights in MEF are time-varying matrices that are in the information set one period earlier and depend on the conditional variance of the martingale increment and the conditional mean of the parameter derivatives (see Christensen and Sørensen, 2008). In the models we consider, these optimal weights can be analytically derived and depend on the structural model parameters. MEF exactly identifies all structural parameters of the system, through its set-up and by having both the weights and martingale increments depend on the structural parameters. We contrast this estimation approach to standard GMM, which we prove is inefficient relative to the MEF estimator.[7]

We extend the MEF approach in two directions. First, we further develop the mixed-frequency nature of our set-up. The baseline MEF approach utilizes financial market data at a daily level to develop proxies at the lower frequency (monthly or quarterly) when all the macro variables are observed at the latter. Our extended Mixed-Frequency MEF (MF-MEF) approach allows for the macro variables themselves to be observed at different frequencies. We stipulate the model at a high frequency, monthly in our applications, and use the other variables at the frequencies at which they occur. For example, we consider the monthly consumption series together with the quarterly GDP series. We implement the approach by considering the actual quarterly output in the month it becomes available and model-based predictions in the two months within the quarter where output is unavailable. Monthly consumption and quarterly GDP are combined with the monthly proxies based on the daily financial variable. Using mixed-frequency data for estimation has recently received considerable attention in the literature, see, e.g., Ghysels, Sinko, and Valkanov (2007), Andreou, Ghysels, and Kourtellos (2010, 2013), and Schorfheide and Song (2014). Our MF-MEF approach provides a structural approach for mixed-frequency estimation.

Second, we extend the MEF set-up to allow for latent variables. In the Simulated MEF (SMEF) approach we simulate the latent variables using the process implied by the model. The simulated paths are used to obtain the conditional expectations that are used in the MEF approach. Model-consistent proxies are used to keep the simulated paths from becoming too dispersed over time. In our implementation we consider three cases: (i) An unobserved interest rate, (ii) a regime-switching spot rate volatility process, and (iii) stochastic volatility, but the approach can be generalized to include other latent variables, such as expected

[7]We consider the version of GMM common in practice, based on the same (small) number of conditional moments as MEF, and expanding these using instruments to get sufficiently many moments. We also discuss the relation between this standard GMM procedure, MEF, and the GMM literature on efficiency bounds and optimal instruments. In addition, we compare to regression-based two-step procedures of first estimating reduced form parameters, then obtaining structural parameters by minimum distance. These procedures are inefficient relative to MEF, too, and do not fully correct for endogeneity in the system.

inflation, (stochastic) discount rates, etc.

Though there is a tradition in macroeconomics to model financial frictions, we illustrate our approach for a simple DSGE model where the equilibrium dynamics are available in terms of observable quantities, namely, consumption, output, and interest rates. It is important to study the properties of our estimation methods in simple continuous-time models before addressing more elaborate models at the vanguard of the DSGE literature, including financial intermediation and frictions.[8] Indeed, financial frictions may help increase the persistence of temporary shocks, and lead to amplification and propagation of shocks. One shortcoming of our specification is that it is not explaining the dynamics of high-frequency financial data, in particular, the effects from the macroeconomy on the dynamics of the interest rate. In our macro-finance model, financial shocks do translate from financial variables to the real economy (see Jermann and Quadrini, 2012). Nonetheless, the channels through which financial frictions affect the macroeconomy and through which shocks are propagated to financial variables are missing.[9]

We consider logarithmic preferences together with a linear technology as an important benchmark case because it allows for an analytical solution of the continuous-time model. Moreover, we have some choice in modeling the interest rate dynamics. Specifications of this kind date back to Cox, Ingersoll, and Ross (1985a), and are frequently used in macro-finance models. Since the MEF approach is not limited to analytical solutions, but applicable to cases where the researcher is provided with a solution in the form of policy functions, our illustrating example can be used as a point of reference for exploring broader classes of dynamic general equilibrium models.

We apply our model to both simulated and empirical data on production, consumption, and interest rates. Our Monte Carlo study examines the properties of our estimation approaches in 1,000 simulated data sets of 25 years each for both monthly and quarterly macro data, along with daily financial market data, roughly in line with the availability of empirical figures. The results show that the GMM and MEF approaches generally are able to accurately estimate the parameters of the (correctly specified) model, and that the interest rate data help identifying the structural parameters. When differences appear, MEF is more precise than GMM, and in some cases identifies more parameters, whereas both approaches are preferred over alternatives based on regression and minimum distance.

[8]The literature on financial frictions emerged from the seminal contributions by Bernanke and Gertler (1989); Kiyotaki and Moore (1997); Bernanke, Gertler, and Gilchrist (1999). Recent contributions include Liu, Wang, and Zha (2013) and Christiano, Motto, and Rostagno (2014).

[9]For example, as in Brunnermeier and Sannikov (2014), due to occasionally binding constraints the effects of small shocks may depend on the state of the economy, say, normal times versus crisis episodes. Accounting for such regime dependence calls for our MEF extensions to stochastic volatility and/or regime-switching.

Our empirical application to 30 years of U.S. data shows that it is possible to estimate our macro-finance model using macro and high-frequency financial data in combination, although this simple model is likely to be misspecified. The results imply a long run mean of the interest rate around 10% with a 1.5% volatility annually and only weak mean reversion. Comparing our simulated results to empirical estimates and/or the estimated interest rate, when treating it as a latent variable, indicates potential misspecification of the simple model. Nonetheless, due to its simplicity and tractability, the AK-Vasicek model specification should provide a benchmark for future research.

The paper proceeds as follows. Section 2 summarizes the macroeconomic theory and solution techniques, and provides a comparison to the discrete-time model. Section 3 presents the estimation strategies. Sections 4 and 5 provide Monte Carlo evidence on small sample properties of our estimation strategies and report empirical estimates. Section 6 concludes. Technical derivations and proofs of main results are deferred to the appendix, Section 7. Further derivations and results are available in an accompanying web appendix.

2 Framework

Our model is cast in continuous time (Eaton, 1981; Cox, Ingersoll, and Ross, 1985a). This allows the application of Itô's calculus, and in some cases we can solve the model analytically to obtain closed-form expressions which facilitates statistical inference. However, as much research in the DSGE literature is using discrete-time models, we also sketch how to adapt our framework to discrete time.

2.1 The Macro-Finance model

Production possibilities. At each point in time, certain amounts of capital, labor, and factor productivity are available in the economy, and these are combined to produce output. The production function is a constant returns to scale technology subject to regularity conditions (see Chang, 1988),

$$Y_t = A_t F(K_t, L), \tag{1}$$

where K_t is the aggregate capital stock, L is the constant population size, and A_t is total factor productivity (TFP), in turn driven by a standard Brownian motion B_t,

$$dA_t = \mu(A_t)dt + \eta(A_t)dB_t, \tag{2}$$

with $\mu(A_t)$ and $\eta(A_t)$ generic drift and volatility functions satisfying regularity conditions.[10] The capital stock increases if gross investment I_t exceeds capital depreciation,

$$dK_t = (I_t - \delta K_t)dt + \sigma K_t dZ_t, \tag{3}$$

where δ denotes the mean and σ the volatility of the stochastic depreciation rate, driven by another standard Brownian motion Z_t.

Equilibrium properties. In equilibrium, factors of production are rewarded with marginal products $r_t = Y_K$ and $w_t = Y_L$, subscripts K and L indicating derivatives with respect to K_t and L, and the goods market clears, $Y_t = C_t + I_t$. By an application of Itô's formula (e.g., Protter, 2004; Sennewald, 2007), the technology in (2), capital accumulation in (3), and market clearing condition together imply that output evolves according to

$$\begin{aligned} dY_t &= Y_A dA_t + Y_K dK_t + \tfrac{1}{2} Y_{KK} \sigma^2 K_t^2 dt \\ &= (\mu(A_t) Y_A + (I_t - \delta K_t) Y_K + \tfrac{1}{2} Y_{KK} \sigma^2 K_t^2) dt + Y_A \eta(A_t) dB_t + \sigma Y_K K_t dZ_t. \end{aligned} \tag{4}$$

This corresponds to equation (1) in Cox, Ingersoll, and Ross (1985a) (henceforth CIR), where $I_t - \delta K_t$ is the amount of the output good allocated to the production process. In general, Y_t can be a nonlinear activity, determined by the output elasticity of capital.[11]

Preferences. We consider an economy with a single consumer, which we interpret as a representative "stand-in" for a large number of identical consumers. The consumer maximizes expected additively separable discounted life-time utility given by

$$U_0 \equiv E_0 \int_0^\infty e^{-\rho t} u(C_t, A_t) dt, \quad u_C > 0, \ u_{CC} < 0, \tag{5}$$

subject to

$$dK_t = ((r_t - \delta) K_t + w_t L - C_t) dt + \sigma K_t dZ_t, \tag{6}$$

where ρ is the subjective rate of time preference, r_t is the rental rate of capital, and w_t is the labor wage rate. We do not consider financial claims, which can be thought of as being in zero net supply. The paths of factor rewards are taken as given by the representative consumer. The generic utility flow function specification $u(C_t, A_t)$ allows the possibility that technology enters as an argument. This may represent a quest for technology and is included for comparability with CIR.

[10] We assume that $E(A_t) = A \in \mathbb{R}_+$ exists, and that the integral describing life-time utility in (5) below is bounded, so that the value function is well-defined.

[11] Unless we consider a nonlinear production process, our model is formally included in the CIR economy. We are not aware of any other paper estimating the model's structural parameters using macro and financial data.

2.2 The Euler equation

The relevant state variables are capital and technology, (K_t, A_t). For given initial states, the value of the optimal program is

$$V(K_0, A_0) = \max_{\{C_t\}_{t=0}^{\infty}} U_0 \quad s.t. \quad (6) \quad \text{and} \quad (2), \tag{7}$$

i.e., the present value of expected utility along the optimal program.[12] It is shown in Appendix 7.1 that the first-order condition for the problem is

$$u_C(C_t, A_t) = V_K(K_t, A_t), \tag{8}$$

for any $t \in [0, \infty)$ (subscripts C and K denoting derivatives), and this allows writing consumption as a function of the state variables, $C_t = C(K_t, A_t)$. The *Euler equation* is

$$\begin{aligned} \frac{du_C}{u_C} &= (\rho - (r_t - \delta))dt - \frac{u_{CC}(C_t, A_t)}{u_C(C_t, A_t)} C_K \sigma^2 K_t dt + \frac{u_{CC}(C_t, A_t)}{u_C(C_t, A_t)} C_A \eta(A_t) dB_t \\ &\quad + \frac{u_{CA}(C_t, A_t)}{u_C(C_t, A_t)} \eta(A_t) dB_t + \frac{u_{CC}(C_t, A_t)}{u_C(C_t, A_t)} C_K \sigma K_t dZ_t. \end{aligned} \tag{9}$$

Economically, this gives the pricing kernel in the economy. Hence, we may use the Euler equation (9) to shed light on how the rate of return on the physical asset is linked to any risk-free security (cf. Posch, 2011). For this purpose we apply the conditional expectation and rewrite terms to arrive at

$$\underbrace{\rho - \frac{1}{dt} E_t \left[\frac{du_C}{u_C} \right]}_{\text{cost of forgone consumption}} = \underbrace{r_t - \delta + \frac{u_{CC}(C_t, A_t)}{u_C(C_t, A_t)} C_K \sigma^2 K_t}_{\text{certainty equivalent rate of return}} \equiv r_t^f. \tag{10}$$

Optimal behavior implies that the cost of forgone consumption on the left-hand side must equal the certainty equivalent rate of return r_t^f, corresponding to the rate on the instantaneously risk-free asset on the right-hand side of the equation.[13]

Moreover, the Euler equation determines the optimal consumption path. In the following, we restrict attention to the case $u(C_t, A_t) = u(C_t)$. Using the inverse marginal utility function, we obtain the path for consumption,

$$\begin{aligned} dC_t &= \frac{u'(C_t)}{u''(C_t)} (\rho - (r_t - \delta))dt - \sigma^2 C_K K_t dt - \tfrac{1}{2}(C_A^2 \eta(A_t)^2 + C_K^2 \sigma^2 K_t^2) \frac{u'''(C_t)}{u''(C_t)} dt \\ &\quad + C_A \eta(A_t) dB_t + C_K \sigma K_t dZ_t, \end{aligned} \tag{11}$$

where $u' > 0$ and $u'' < 0$ (strict concavity of preferences).

[12] Christensen and Kiefer (2009) is a textbook reference on economic modeling and inference using dynamic programming models.

[13] Note that $-u_{CC}(C_t, A_t) C_t / u_C(C_t, A_t)$ measures the degree of relative risk aversion.

2.3 Equilibrium dynamics of the economy

Applying the logarithm to the variables of the stochastic differentials (6), (4), and (11) the equilibrium dynamics of the economy may be summarized by the instantaneous growth rates

$$\begin{aligned}
d\ln C_t &= \left(\frac{u'(C_t)(\rho - r_t + \delta)}{u''(C_t)C_t} - \frac{C_K K_t \sigma^2}{C_t} - \tfrac{1}{2}\frac{C_A^2 \eta(A_t)^2 + C_K^2 \sigma^2 K_t^2}{C_t^2}\frac{u'''(C_t)C_t + u''(C_t)}{u''(C_t)}\right)dt \\
&\quad + C_A \eta(A_t)/C_t dB_t + C_K \sigma K_t / C_t dZ_t, \\
d\ln Y_t &= \left(\frac{\mu(A_t)}{A_t} + \left(\frac{Y_t - C_t}{K_t} - \delta\right)\frac{K_t Y_K}{Y_t} + \tfrac{1}{2}\sigma^2 \frac{K_t^2 Y_{KK}}{Y_t}\right)dt - \tfrac{1}{2}\frac{Y_A^2 \eta(A_t)^2 + \sigma^2 Y_K^2 K_t^2}{Y_t^2}dt \\
&\quad + Y_A \eta(A_t)/Y_t dB_t + \sigma Y_K K_t / Y_t dZ_t, \\
d\ln K_t &= (r_t - \delta + w_t/K_t - C_t/K_t - \tfrac{1}{2}\sigma^2)dt + \sigma dZ_t.
\end{aligned}$$

If all variables C_t, Y_t, and K_t along with TFP A_t were observed, estimation could be based directly on this system and (2). While consumption and income are standard variables in most macro studies, capital and technology are notoriously problematic, due to the risk of mismeasurement. This is where we propose using financial variables in a unified macro-finance framework, instead. The idea is to use model-based equilibrium conditions to identify latent state variables using financial data. Thus, suppose that an interest rate r_t is identified, either directly in the data, or in the form of an equilibrium no-arbitrage condition such as r_t^f in (10), along with C_t and Y_t.[14] We consider systems of stochastic differential equations that can be used for estimation in this case, based on time series data on (C_t, Y_t, r_t), by recasting the equilibrium dynamics in terms of this triple.

2.4 AK-Vasicek model with logarithmic preferences

In this section we consider an economy with technology given by $Y_t = A_t K_t$, also known as the AK model (this includes the technology in Brunnermeier and Sannikov, 2014), and assume preferences of the type $u(C_t) = \ln C_t$. Our specification is interpreted as a parsimonious description which allows us to study the macro-finance links in production economies with optimizing agents. It is used as a benchmark and illustrates the issues and advantages of our approach by making use of an analytical solution. With these assumptions, $A_t = Y_K = r_t$ and $K_t = Y_t/A_t = Y_t/r_t$, so the two relevant state variables (A_t, K_t) are expressed as known functions of the observable variables (Y_t, r_t). In this case we have $w_t = Y_L = 0$, such that

[14] One caveat is that some variables are observed as an integral over an interval (a flow) rather than at a point in time (a stock; Harvey and Stock, 1989). We approximate a flow variable, e.g., $Y_t\Delta$ at time t, by the integral $\int_{t-\Delta}^{t} Y_s ds$. Observed growth rates of flow variables therefore correspond to $\ln Y_t - \ln Y_{t-\Delta}$.

the equilibrium dynamics can be summarized as

$$\begin{aligned} d\ln C_t &= \left(r_t - \rho - \delta - \tfrac{1}{2}\sigma^2\right) dt + \sigma dZ_t, & \text{(12a)} \\ d\ln Y_t &= \left(\mu(r_t)/r_t + r_t - \rho - \delta - \tfrac{1}{2}\eta(r_t)^2/r_t^2 - \tfrac{1}{2}\sigma^2\right) dt \\ &\quad + \eta(r_t)/r_t dB_t + \sigma dZ_t, & \text{(12b)} \\ dr_t &= \mu(r_t)dt + \eta(r_t)dB_t. & \text{(12c)} \end{aligned}$$

In more general models, the consumption function is non-homogeneous with respect to TFP A_t and capital K_t (or wealth, here the output-TFP ratio).[15] The functions $\mu(\cdot)$ and $\eta(\cdot)$ are chosen such that boundedness conditions are met (cf. Posch, 2009). Recall that in the AK framework, the interest rate (rental rate of capital) dynamics purely reflect TFP dynamics. In the tradition of the finance literature, we illustrate the estimation of the model with the interest rate governed by a Vasicek specification (henceforth the AK-Vasicek model).

The Vasicek (1977) mean-reverting specification for the rental rate of physical capital is $\mu(r_t) = \kappa(\gamma - r_t)$ and $\eta(r_t) = \eta$, where $\kappa > 0$ is the speed and γ the target rate of mean reversion, and η the constant volatility. In this case, the equilibrium dynamics are

$$\begin{aligned} d\ln C_t &= \left(r_t - \rho - \delta - \tfrac{1}{2}\sigma^2\right) dt + \sigma dZ_t, & \text{(13a)} \\ d\ln Y_t &= \left(\kappa\gamma/r_t - \tfrac{1}{2}\eta^2/r_t^2 + r_t - \kappa - \rho - \delta - \tfrac{1}{2}\sigma^2\right) dt + \eta/r_t dB_t + \sigma dZ_t, & \text{(13b)} \\ dr_t &= \kappa(\gamma - r_t)dt + \eta dB_t. & \text{(13c)} \end{aligned}$$

Alternative Markov diffusion specifications of the interest rate process, such as the Cox, Ingersoll, and Ross (1985b) square root process or others (see, e.g., Aït-Sahalia, 1996), can be implemented and the system estimated along the lines developed below. The analytical solution does not depend on this particular choice.

In this AK-Vasicek model, the relation between the risk-free rate and the rental rate of capital in (10) is given by

$$r_t^f = r_t - \delta - \sigma^2. \tag{14}$$

This result is quite intuitive if we recall that consumption is proportional to the capital stock, such that $C_K K_t = C_t$. Thus, the term $u_{CC}(C_t, A_t)C_K K_t / u_C(C_t, A_t)$ in (10) equals minus the coefficient of relative risk aversion, of unit magnitude for logarithmic preferences.[16]

Economically, (14) is the equilibrium asset pricing relationship, prescribing that the rate of return to any riskless financial asset r_t^f equal capital rewards r_t net of the rate of depreciation δ and the risk premium associated with holding the physical asset. The equation

[15] In our benchmark case optimal consumption is linear in the capital stock, $C_t = \rho K_t$ (cf. Appendix 7.1).

[16] Note that δ would also capture a constant level of inflation. This assumption, however, neglects inflation dynamics. There is a number of ways to overcome this, which is part of our research agenda. Since the focus of the present paper is methodological, we leave a thorough examination for future research.

sheds light on the use of empirical data for estimating the macro-finance model: Rather than observing the rental rate of capital r_t (or TFP, or the capital stock), it is typically easier to relate this to an observable riskless rate r_t^f, or a close proxy, such as the 3-month interest rate (cf. Chapman, Long, and Pearson, 1999).

2.5 Discrete-time formulation

In order to accommodate the discrete-time nature of the data, we integrate over $s \geq t$, employing exact solutions whenever possible to obtain an exact discrete-time analog in terms of (observable) variables. In what follows, we treat the triple of variables (C_t, Y_t, r_t^f) as being observable.[17] Specifically, we use the risk-free rate r_t^f rather than a direct measure of the rental rate of capital r_t as financial data.[18] This (more realistic) assumption implies that r_t is linked to the data, but may also depend on parameters. Using the system of differential equations (13) with $s - t$ fixed at Δ together with the asset-pricing condition (14) yields

$$\ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^{t} r_v^f dv = -\left(\rho - \tfrac{1}{2}\sigma^2\right)\Delta + \varepsilon_{C,t}, \tag{15a}$$

$$\begin{aligned}\ln(Y_t/Y_{t-\Delta}) - \int_{t-\Delta}^{t} r_v^f dv &= \kappa\gamma \int_{t-\Delta}^{t} 1/(r_v^f + \delta + \sigma^2)dv - \tfrac{1}{2}\eta^2 \int_{t-\Delta}^{t} 1/(r_v^f + \delta + \sigma^2)^2 dv \\ &\quad - \left(\kappa + \rho - \tfrac{1}{2}\sigma^2\right)\Delta + \varepsilon_{Y,t},\end{aligned} \tag{15b}$$

$$r_t^f = e^{-\kappa\Delta} r_{t-\Delta}^f + (1 - e^{-\kappa\Delta})(\gamma - \delta - \sigma^2) + \varepsilon_{r,t}, \tag{15c}$$

in which we define martingale increments by

$$\varepsilon_{C,t} \equiv \sigma(Z_t - Z_{t-\Delta}), \tag{16a}$$

$$\varepsilon_{Y,t} \equiv \int_{t-\Delta}^{t} \eta/(r_v^f + \delta + \sigma^2)dB_v + \sigma(Z_t - Z_{t-\Delta}), \tag{16b}$$

$$\varepsilon_{r,t} \equiv \eta e^{-\kappa\Delta} \int_{t-\Delta}^{t} e^{\kappa(v-(t-\Delta))} dB_v. \tag{16c}$$

The system (15) of three equations forms the basis of our empirical specifications. At the same time, it illustrates the main ideas underlying our approach. First, our analysis delivers the explicit functional forms of the relations among observables. Second, the availability of interest rate data at higher frequency (say, daily) than consumption and production (monthly or quarterly) allows precise approximation of the ordinary (although not the stochastic) integrals involving the interest rate by summation over days. In our applications, we approximate the integrals by Riemann sums of the type $\int_{t-\Delta}^{t} g(r_v^f)dv \approx \Delta \sum_{i=1}^{P} g(r_{t-\Delta+i\Delta/P}^f)/P$, where

[17] In an extension below, we consider the case with latent interest rates.

[18] We use daily data on the 3-month interest rate as a proxy for the risk-free rate (cf. Chapman, Long, and Pearson, 1999), along with aggregate consumption and output at lower frequencies.

$g(\cdot)$ is a smooth function of $r^f_{t-\Delta+i\Delta/P}$, the prevailing interest rate on a risk-free security on day i in the period between $t-\Delta$ and t, and P is the number of days in the period.[19] Third, the structural parameters enter into the coefficients on the terms involving interest rates, thus showing that financial data may serve to identify parameters of interest.

If we directly observed $r_t = r^f_t + \delta + \sigma^2$, the system would in fact be linear in a set of coefficients, say β, on right-hand side variables of the type 1, $\int_{t-\Delta}^t 1/r_v dv$, $\int_{t-\Delta}^t 1/r_v^2 dv$, and $r_{t-\Delta}$. These β-coefficients play the role of reduced-form parameters, and are in turn known functions of the structural parameters, namely, $(\kappa, \gamma, \eta, \rho, \delta, \sigma)^\top$. This intriguing result may be exploited even when r^f_t is observed, instead, such that r_t depends on parameters: We may replace the unobserved time series r_t with any reasonable proxy $\hat{r}_t$ from (14), using particular values $\delta = \delta_0$ and $\sigma = \sigma_0$. Technically, given the values for δ and σ, the rental rate of capital is uniquely identified from the risk-free rate,

$$\hat{r}_t = r^f_t + \delta_0 + \sigma_0^2. \tag{17}$$

This step allows using regression-based estimation methods (a textbook reference is Canova, 2007, see also web appendix Section A on linear regression methods specifically for our problem, and Andreasen and Christensen (2015) for general sequential nonlinear regression methods), with Riemann sums based on $\hat{r}_t$ as feasible regressors approximating the integrals involving r_t, but this substitution is not required for the GMM and MEF methods presented in Section 3 below, as δ and σ entering the right-hand side variables of (15) in a nonlinear fashion is unproblematic for these.

We have some choice in turning system (15) into an empirical specification. We specify a system of three regression equations for equidistant observations (we use $\Delta = 1/12$ for monthly data, $\Delta = 1/4$ for quarterly data). Given the higher (say, daily) frequency of financial data, an alternative would be to start out with separate estimation of the third equation, but the full system is likely closer to that required for more complicated models (e.g., if macro variables enter in the interest rate equation). In any case, the high-frequency property of the interest rate data is exploited in the approximation of the integrals as Riemann sums.

It is important to understand that in order to get to the empirical specification we kept the full nonlinear structure of the model, without employing any approximation to the solution of the economic model and/or to the equilibrium dynamics. This is in contrast to the traditional discrete-time approach which we briefly discuss next.

[19] For notational convenience, we write P as a constant, but in our empirical approach we use the actual number of days in the period (month or quarter).

2.6 Comparison with discrete-time models

Here, we follow the tradition in the business cycle literature by writing our dynamic equilibrium model in discrete time for a given frequency, which is typically aligned with the observation frequency of macro variables. Thus, it is useful to sketch what our approach in discrete time would be:

1. We would write down the equilibrium conditions of the discrete-time model. The main difference is that no analytical solution is available (it is not possible to eliminate the expectation operator from these conditions). This is not problematic but should be kept in mind when comparing the empirical specifications of the model.[20] In our case, the equilibrium dynamics of the model can be summarized as

$$C_t^{-1} = \tilde{\beta} E_t \left[(1-\tilde{\delta}+\tilde{r}_{t+1}) C_{t+1}^{-1} \right], \tag{18a}$$

$$\begin{aligned} Y_{t+1} = {} & Y_t + (\tilde{r}_t - \tilde{\delta})Y_t - \tilde{r}_t C_t + \tilde{\kappa}(\tilde{\gamma} - \tilde{r}_t)Y_t/\tilde{r}_t + \tilde{\eta} Y_t/\tilde{r}_t \epsilon_{A,t+1} + \tilde{\sigma} Y_t \epsilon_{K,t+1} \\ & + ((\tilde{r}_t - \tilde{\delta})Y_t/\tilde{r}_t - C_t + \tilde{\sigma} Y_t/\tilde{r}_t \epsilon_{K,t+1})\left(\tilde{\kappa}(\tilde{\gamma} - \tilde{r}_t) + \tilde{\eta}\epsilon_{A,t+1}\right), \end{aligned} \tag{18b}$$

$$\tilde{r}_{t+1} = \tilde{r}_t + \tilde{\kappa}(\tilde{\gamma} - \tilde{r}_t) + \tilde{\eta}\epsilon_{A,t+1}, \tag{18c}$$

where $\tilde{r}_t$ denotes the periodic interest rate (e.g., quarter-to quarter for quarterly model), $\tilde{\kappa}, \tilde{\gamma}, \tilde{\eta}, \tilde{\beta}, \tilde{\delta}$, and $\tilde{\sigma}$ are the structural parameters for the given frequency, and $\epsilon_{A,t}$ and $\epsilon_{K,t}$ are the shocks corresponding to the differentials dB_t and dZ_t, respectively.[21]

2. One way of proceeding is to log-linearize the economic model to arrive at a system of equilibrium dynamics that can be used for estimation. This yields

$$\ln(C_t/C_{t-1}) = \ln\tilde{\beta} + \tilde{r}^f_{t-1} + \mathbb{C}_0 + \tilde{\varepsilon}_{C,t}, \tag{19a}$$

$$\ln(Y_t/Y_{t-1}) = \ln\tilde{\beta} + \tilde{r}^f_{t-1} + \mathbb{C}_0 - \mathbb{C}_2\left(\tilde{r}^f_{t-1} - \mathbb{C}_1\right) + \tilde{\varepsilon}_{Y,t}, \tag{19b}$$

$$\tilde{r}^f_t = (1-\tilde{\kappa})\tilde{r}^f_{t-1} + \tilde{\kappa}\mathbb{C}_1 + \tilde{\varepsilon}_{r,t}, \tag{19c}$$

where $\mathbb{C}_0 \equiv \frac{1}{2}((\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2)/((1-\tilde{\delta}+\tilde{\gamma})^2)$, $\mathbb{C}_1 \equiv \tilde{\gamma} - \tilde{\delta} - \mathbb{C}_0$, $\mathbb{C}_2 \equiv (1-\tilde{\delta})\tilde{\kappa}/((1-\tilde{\kappa})\tilde{\gamma})$,

$$\tilde{\varepsilon}_{C,t} \equiv \frac{\tilde{\sigma}}{\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})}\epsilon_{K,t} + \frac{\tilde{\eta}}{1-\tilde{\delta}+\tilde{\gamma}}\epsilon_{A,t}, \tag{20a}$$

$$\tilde{\varepsilon}_{Y,t} \equiv \frac{\tilde{\sigma}}{\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})}\epsilon_{K,t} + \frac{\tilde{\eta}}{\tilde{\gamma}}\epsilon_{A,t}, \tag{20b}$$

$$\tilde{\varepsilon}_{r,t} \equiv \frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}\tilde{\eta}\epsilon_{A,t}. \tag{20c}$$

[20] The interested reader is referred to the web appendix (cf. Section B) where we present the discrete-time model, the equilibrium conditions, the log-linear solution, and the equivalent empirical specification.

[21] We use the mapping $\tilde{\kappa} = 1 - e^{-\Delta\kappa}$, $\tilde{\gamma} = \Delta\gamma$, $\tilde{\eta} = \Delta\eta\sqrt{(1-e^{-2\kappa\Delta})/(2\kappa)}$, $\tilde{\beta} = e^{-\Delta\rho}$, $\tilde{\delta} = 1 - e^{-\Delta\delta}$, and $\tilde{\sigma} = \Delta^{1/2}\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})\sigma$, in which $\Delta = 1/12$ for the monthly model, and $\Delta = 1/4$ for the quarterly model.

Here, $\tilde{r}_t^f$ denotes the periodic risk-free interest rate.

At the expense of introducing approximation error due to log-linearization, the system (19) is now linear in a set of reduced-form parameters, which in turn are known functions of the structural parameters.[22] This system can be directly compared to system (15) of the continuous-time model. While the analytical solution in the continuous-time version implies that consumption should not respond to shocks to the interest rate, the discrete-time interest residuals (20c) do enter the approximate solution for the consumption residuals (20a). Another feature of the log-linear approximation is that no endogenous variable appears in the error term. This is in error, as is known from the exact solution in continuous time. The path of the interest rate during the course of the month enters in the exact output residual (16b).

3 Estimation: The MEF approach

In this section, we describe how to estimate the equilibrium system (15) using macro and financial data. In fact, the system permits the use of standard regression-based methods. We consider OLS, SUR (to account for cross-equation correlation), instrumental variables (IV, to address endogeneity issues), and a feasible combination which we label FGLS-SUR-IV and which appears to be novel.[23] Although these methods share the advantages of being easy to understand and implement, none of them fully corrects for endogeneity in the structural model. The endogeneity stems from the right-hand side variables in (15) including two integrals involving the evolution of the auxiliary variable in (17) from $t-\Delta$ through t and so being correlated with both $\varepsilon_{r,t}$ and $\varepsilon_{Y,t}$. A standard IV approach is to consider first-stage regressions of the right-hand side variables on lags and an intercept, then use fitted values to perform the main regressions. However, the lagged values of the relevant integrals involving the auxiliary variable $\hat{r}_s$, $t-2\Delta \leq s \leq t-\Delta$, may correlate with $\hat{r}_{t-\Delta}$, and hence with $\varepsilon_{Y,t}$ from (16b), although presumably less than without lagging (this is the idea of the instrumentation). Any such correlation between the error terms and the right-hand side variables (even when using fitted values) indicates that part of the endogeneity issue remains.

Hence, we show how structural parameters may be directly estimated using the martingale estimating function (MEF) approach which is efficient and fully corrects the endogeneity problem. We also show how this approach relates to the generalized method of moments

[22] The log-linear solution implies $\ln(C_t/K_t) = \ln(1-\tilde{\beta}) + \ln(1-\tilde{\delta}+\tilde{\gamma}) + (\tilde{r}_t - \tilde{\gamma})/(1-\tilde{\delta}+\tilde{\gamma})$.

[23] We show in the web appendix how to view our system as a regression model, how to estimate the reduced-form parameters by these regression-based methods, and how to obtain the structural parameters using minimum distance (cf. Section A.1 in the web appendix).

(GMM). The regression-based approaches serve as useful benchmarks and starting values for the MEF estimator, which is based on the theory of optimal estimators.

Let ϕ denote the parameter vector, and $m_t = m_t(\phi)$ the vector of martingale increments generated by the model, expressed in terms of data and parameters. In this section, we study the MEF and GMM estimators from a general point of view, but to fix ideas, it is useful to recall that our specific application has $m_t = \varepsilon_t = (\varepsilon_{C,t}, \varepsilon_{Y,t}, \varepsilon_{r,t})^\top$ from (16a)-(16c), and $\phi = (\kappa, \gamma, \eta, \rho, \delta, \sigma)^\top$. Clearly, m_t is a martingale difference sequence, and from system (15) we have that in terms of data and parameters

$$m_t = \begin{pmatrix} \ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^t r_v^f dv + \left(\rho - \frac{1}{2}\sigma^2\right)\Delta \\ \ln(Y_t/Y_{t-\Delta}) - \int_{t-\Delta}^t r_v^f dv + \left(\kappa + \rho - \frac{1}{2}\sigma^2\right)\Delta - \kappa\gamma \int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2) dv \\ \qquad + \frac{1}{2}\eta^2 \int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)^2 dv \\ r_t^f - (1 - e^{-\kappa\Delta})(\gamma - \delta - \sigma^2) - e^{-\kappa\Delta} r_{t-\Delta}^f \end{pmatrix}, \tag{21}$$

where the integrals are approximated by Riemann sums over days between $t-\Delta$ and t. More general versions of the model below give rise to other m_t, some with higher dimension.

3.1 From GMM to MEF

The MEF method differs from the generalized method of moments (GMM) of Hansen (1982). The relevant theory for optimal estimators is based on Godambe and Heyde (1987), and the dynamic case is treated in Christensen and Sørensen (2008). It is shown that MEF is at least as efficient as GMM – indeed, strictly more efficient except in the special case where the two estimators coincide. Hence, it is instructive to start with the well-known GMM, then show how to modify this appropriately, to see how the MEF method comes about.

Since m_t is a martingale difference sequence, we have the conditional moment restrictions $E_{t-\Delta}(m_t) = 0$. The standard GMM approach is to consider instruments, say z_t, belonging to the information set and hence known at time $t-\Delta$, so that $E_{t-\Delta}(z_t \otimes m_t) = z_t \otimes E_{t-\Delta}(m_t) = 0$, where $\otimes$ is the Kronecker product. For example, the instruments could be lagged right-hand side variables, $z_t = (1, \int_{t-2\Delta}^{t-\Delta} 1/(r_v^f + \delta + \sigma^2) dv, \int_{t-2\Delta}^{t-\Delta} 1/(r_v^f + \delta + \sigma)^2 dv, r_{t-2\Delta}^f)^\top$, since these are all in the information set at $t-\Delta$. In particular, it presents no new issue, neither for GMM nor MEF, that the instrumental variables depend not only on data, but also on parameters, $z_t = z_t(\phi)$. Defining $h_t = h_t(\phi) = z_t \otimes m_t$, we have that h_t is of dimension $\dim h = \dim z \cdot \dim m$, or 12 in the AK-Vasicek model with logarithmic utility, where $\dim m = 3$ and $\dim z = 4$ in the example. To construct the GMM estimator, let for notational convenience $\Delta = 1$ and define

$$H_T = \sum_{t=1}^{T} h_t, \tag{22}$$

so that H_T/T is the sample average of $\{h_t\}$. Evidently, H_T is a martingale at the true value of the parameter ϕ, i.e., $E_T(H_{T+1}) = H_T$, because $E_T(h_{T+1}) = z_{T+1} \otimes E_T\left(m_{T+1}\right) = 0$.[24] Since the unconditional expectation $E(h_t) = 0$, it would be natural to choose the estimator for ϕ so as to equate the sample analogue H_T/T of $E(h_t)$ to zero. Typically, $\dim h > \dim \phi$, so it is not possible to solve the equation $H_T = 0$ exactly. Instead, the GMM estimator is defined as the minimizer of the squared norm $H_T(\phi)^\top W H_T(\phi)$, where W is a weight matrix. Optimal GMM is obtained by using the identity matrix $I_{\dim h}$ for W in a first step minimization, then using the resulting estimator, say $\hat{\phi}_0$, to calculate an estimate of $Var(H_T)^{-1}$ that is used for W in the second step minimization. Frequently $W = (\sum_t h_t(\hat{\phi}_0)h_t(\hat{\phi}_0)^\top)^{-1}$ is used for the second step.[25] Sometimes, a mean adjustment and/or a Newey and West (1987) correction is used, the latter for robustness against serial correlation, but both are unnecessary under the null that m_t and hence h_t is a martingale difference sequence.

The critical features of GMM that leave room for improvement and hence the MEF approach are now evident: First, in GMM, the instruments z_t enter in the form of a vector, whereas MEF uses a matrix. Secondly, in GMM, the dimension of H_T is the same as or greater than the number of parameters, whereas MEF specifically uses the same number of estimating equations and parameters. In short, the optimal estimating equations (those of the MEF method) are based on matrix-valued rather than vector-valued instruments, and over-identifying restrictions ($\dim h > \dim \phi$) are unnecessary for efficiency.

To develop these ideas, note that the first order conditions for the minimization in GMM are

$$\frac{\partial H_T(\phi)^\top}{\partial \phi} W H_T(\phi) = 0, \tag{23}$$

a set of $\dim \phi$ equations. Thus, there is the same number of zero conditions as number of parameters in ϕ, as it should be. An estimator that is asymptotically equivalent to GMM may be obtained by solving the $\dim \phi$ equations

$$G \sum_{t=1}^{T} h_t(\phi) = 0, \tag{24}$$

where G is an initial consistent estimate of the $\dim \phi \times \dim h$ matrix $\partial H_T(\phi)^\top / \partial \phi \cdot W$ in (23). For example, G could be based on the first step GMM estimator, just like W, i.e., the

[24] For the general case with $\Delta = 1/N$, say, we may relabel $h_\Delta, h_{2\Delta}, \ldots, h_{S-\Delta}, h_S$ as $h_1, h_2, \ldots, h_{T-1}, h_T$, where $T = NS$ and Δ is the new time unit, then define H_T by (22) without change. Thus, $E_t\left(m_{t+\Delta}\right) = 0$ in the original notation is equivalent to $E_t\left(m_{t+1}\right) = 0$ in the new, and similarly for h_t, i.e., $\Delta = 1$ is w.l.o.g.

[25] In this case, $T \cdot W - Var(H_T/\sqrt{T})^{-1} \to 0$, a.s., and $H_T^\top W H_T = (H_T/\sqrt{T})^\top (T \cdot W)(H_T/\sqrt{T})$ converges in law to χ^2.

system is

$$\left(\sum_t \frac{\partial h_t(\hat{\phi}_0)^\top}{\partial \phi}\right)\left(\sum_t h_t(\hat{\phi}_0)h_t(\hat{\phi}_0)^\top\right)^{-1}\sum_t h_t(\phi) = 0,$$

where ϕ only appears in the last factor. An estimator is obtained by treating $G(\hat{\phi}_0)$ as fixed and finding ϕ that sets the equations exactly equal to zero, and this is asymptotically equivalent to optimal GMM. In essence, this is a way of computing the optimal GMM estimator.

It is now apparent that a more flexible estimation approach obtains by not just solving the equations with a fixed $\dim\phi\times\dim h$ matrix G from the first step (the approach asymptotically equivalent to optimal GMM), but instead allowing a separate $\dim\phi\times\dim h$ matrix each time period, say, g_t, which may depend on data through $t-\Delta$. This is the central idea of the MEF approach. Thus, there are again $\dim\phi$ equations, but they now take the more general form

$$\sum_{t=1}^{T} g_t(\hat{\phi}_0)h_t(\phi) = 0, \tag{25}$$

instead of $G(\hat{\phi}_0)\sum_{t=1}^{T} h_t(\phi) = 0$. Clearly, (25) is a zero-mean martingale for any choice of weight (or instrument) matrices g_t in the information set, since $E_T(g_{T+1}h_{T+1}) = g_{T+1}E_T(h_{T+1}) = 0$. The g_t may also depend on parameters, and here we may again use initial consistent estimates, i.e., all g_t may be calculated after the first step estimation.

The question is how to choose $\{g_t\}$ optimally. If they indeed vary across time, the resulting estimator differs from optimal GMM, since this corresponds to the special case of constant $g_t \equiv G(\hat{\phi}_0)$. In fact, it is unnecessary to expand m_t to h_t by introducing the instruments z_t in $h_t = z_t \otimes m_t$, since if the conditional moment restrictions $E_{t-1}(m_t) = 0$ are used instead of $E_{t-1}(h_t) = 0$, and in fact z_t is needed in the optimal estimator, then z_t will just be part of the optimally chosen g_t in the MEF approach (this fact and the form of the optimal g_t follow from the theorem below). Therefore, we leave the problem involving z_t and define the martingale estimating function

$$M_T = \sum_{t=1}^{T} w_t m_t, \tag{26}$$

clearly a zero-mean martingale for any choice of weight matrices w_t, which may depend on data through $t-1$. This is the case where $h_t = m_t$ (no z_t is used because the optimal estimator is the same with or without z_t, cf. the theorem below) and thus g_t is $\dim\phi\times\dim m$ instead of $\dim\phi\times\dim h$ (this is highlighted by writing w_t instead of g_t). A martingale estimating function (or MEF) is given by specifying w_t as a series of $\dim\phi\times\dim m$ matrices. At the true parameter value, $E(M_T) = 0$, and ϕ is estimated by solving the martingale estimating equation

$$M_T(\phi) = 0. \tag{27}$$

Recall again that the conditions leading to this estimator are

$$E_{t-1}\left(m_t\right) = 0, \tag{28}$$

which we refer to as the conditional moment restrictions based on m_t. Before stating the main results, we define formally the GMM and MEF procedures.

Definition 3.1 *(a) For given conditional moment restrictions based on m_t, a GMM estimator is an estimator obtained as follows: (i) Select a vector of instruments z_t belonging to the information set at $t-1$, and let $h_t = h_t(\phi) = z_t \otimes m_t$, with $\dim z$ sufficiently high for $\dim h = \dim z \cdot \dim m \geq \dim \phi$; (ii) select a positive definite $\dim h \times \dim h$ weight matrix W; (iii) let $H_T = \sum_{t=1}^{T} h_t$; (iv) then the GMM estimator is the minimizer with respect to ϕ of the squared norm $H_T(\phi)^\top W H_T(\phi)$.*

(b) For given conditional moment restrictions and instruments, the optimal GMM estimator is obtained as follows: (i) Use the identity matrix $I_{\dim h}$ for W and compute the associated GMM estimator $\hat{\phi}_0$; (ii) then the optimal GMM estimator is the GMM estimator using $W = (\sum_t h_t(\hat{\phi}_0) h_t(\hat{\phi}_0)^\top)^{-1}$.

(c) For given conditional moment restrictions, an MEF estimator is an estimator obtained as follows: (i) Select matrices of instruments w_t belonging to the information set at $t-1$, of dimension $\dim \phi \times \dim m$; (ii) let $M_T = \sum_{t=1}^{T} w_t m_t$; (iii) then the MEF estimator is the solution with respect to ϕ of the system of $\dim \phi$ equations given by $M_T = 0$.

(d) For given conditional moment restrictions, the optimal MEF estimator is obtained as follows: (i) Let

$$w_t = \psi_t^\top (\Psi_t)^{-1}, \tag{29}$$

where Ψ_t is the conditional variance of the vector martingale increment,

$$\Psi_t = Var_{t-1}(m_t) = E_{t-1}(m_t m_t^\top), \tag{30}$$

and ψ_t the conditional mean of its parameter derivative

$$\psi_t = E_{t-1}\left(\frac{\partial m_t}{\partial \phi^\top}\right); \tag{31}$$

(ii) then the optimal MEF estimator is the MEF estimator using instruments $w_t = \psi_t^\top (\Psi_t)^{-1}$.

The conditioning on information available through $t-1$ (more generally, $t-\Delta$) in (30) and (31) requires integrating out with respect to the evolution of the interest rate appearing in the integrals from $t-\Delta$ through t in (21). This leaves MEF computationally more demanding than GMM and the regression-based approaches, but it does circumvent the endogeneity problem in the DSGE model.

The weights (or instruments) (29) yield the optimal martingale estimating function, across choice of weights w_t. The optimal weights do depend on parameters, i.e., the martingale estimate $\hat{\phi}$ solves a system of the form $\sum_t w_t(\phi)m_t(\phi) = 0$, where the solution accounts for the parameter dependence of both w_t and m_t. Alternatively, an asymptotically equivalent estimator may be obtained by using weights evaluated at initial consistent estimates $\hat{\phi}_0$, e.g., from GMM. In this case, $\hat{\phi}$ is calculated as the solution with respect to ϕ of the system $\sum_t w_t(\hat{\phi}_0)m_t(\phi) = 0$. It is well understood that this does not change the asymptotic properties, and so the proof of the following theorem uses the simpler form of the estimating function, with w_t not explicitly depending on parameters in the derivation.

For convenience and comparison when stating our main results, we include in Theorem 3.2 (a) and (b) the standard properties of GMM, which follow from Hansen (1982). Proof of the remaining items (c) through (g) are provided in Appendix 7.2. The setting is given by Definition 3.1, along with standard regularity conditions used in the GMM literature, in particular, h_t, m_t, and z_{t+1} depend only on data through t, the sample averages $T^{-1}\sum_{t=1}^{T} \partial h_t^\top/\partial\phi$, $T^{-1}\sum_{t=1}^{T} h_t h_t^\top$, and $T^{-1}\sum_{t=1}^{T} \psi_t^\top(\Psi_t)^{-1}\psi_t$ in the statement of the theorem converge either almost surely or in probability to deterministic limits, the latter two limits being invertible matrices, the model is uniquely identified in that $E(M_T)$ is zero only at the true ϕ, and the observed series are stationary and ergodic, which are standard assumptions for interest rates and the growth rates of consumption and output.

Theorem 3.2 *(a) The optimal GMM estmator is consistent and asymptotically normal,*

$$\sqrt{T}(\hat{\phi}_{GMM} - \phi) \to \mathcal{N}(0, V_{GMM}), \tag{32}$$

with asymptotic variance-covariance matrix given by

$$V_{GMM} = \left(E\left(\frac{\partial h_t}{\partial \phi^\top}\right)^\top Var(h_t)^{-1} E\left(\frac{\partial h_t}{\partial \phi^\top}\right)\right)^{-1}, \tag{33}$$

consistently estimated by

$$\hat{V}_{GMM} = \left(\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial h_t^\top}{\partial \phi}\right)\left(\frac{1}{T}\sum_{t=1}^{T} h_t h_t^\top\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial h_t}{\partial \phi^\top}\right)\right)^{-1}. \tag{34}$$

(b) *For given conditional moment restrictions and instruments, the optimal GMM estimator is asymptotically at least as efficient as any other GMM estimator (using different* W*).*

(c) *The optimal MEF estimator is consistent and asymptotically normal,*

$$\sqrt{T}(\hat{\phi} - \phi) \rightarrow \mathcal{N}(0, V_{MEF}), \tag{35}$$

with asymptotic variance-covariance matrix given by

$$V_{MEF} = \left(E(\psi_t^\top (\Psi_t)^{-1} \psi_t)\right)^{-1}, \tag{36}$$

consistently estimated by the inverse sample average

$$\hat{V}_{MEF} = \left(\frac{1}{T} \sum_{t=1}^{T} \psi_t^\top (\Psi_t)^{-1} \psi_t\right)^{-1}. \tag{37}$$

(d) *For given conditional moment restrictions, the optimal MEF estimator is asymptotically at least as efficient as any other MEF estimator (using different* $\{w_t\}$*).*

(e) *For given conditional moment restrictions, the optimal MEF estimator is asymptotically at least as efficient as any GMM estimator,* $V_{MEF} \leq V_{GMM}$*, regardless how the instruments* z_t *are chosen in the GMM. Furthermore, unless the two estimators coincide, the optimal MEF estimator is strictly more efficient than GMM,*

$$V_{MEF} < V_{GMM}. \tag{38}$$

(f) *The event that the optimal MEF estimator and the GMM estimator coincide occurs only in two special cases: (i) The number of parameters equals the number of moment conditions,* $\dim \phi = \dim m$*, and the moment conditions satisfy the condition*

$$m_t(\phi) = \psi_t^\top (\Psi_t)^{-1} m_t(\phi); \tag{39}$$

(ii) The number of parameters exceeds the number of moment conditions, $\dim \phi > \dim m$*, and the product of the conditional expected parameter derivative and conditional variance of the moment conditions has special sparse structure, with*

$$\psi_t^\top (\Psi_t)^{-1} = z_t \otimes I_{\dim m}, \tag{40}$$

where z_t *is the vector of instruments used in GMM, i.e., the GMM is based on* $h_t = z_t \otimes m_t$*, and* $\dim \phi = \dim z \cdot \dim m$*.*

(g) *The optimal MEF estimator based on the conditional moment restrictions* $E_{t-1}(m_t) = 0$ *coincides with the optimal MEF estimator based on the expanded set of moment conditions* $E_{t-1}(h_t) = 0$*, with* $h_t = z_t \otimes m_t$ *and where* z_t *belongs to the information set at* $t-1$.

Proof. cf. Appendix 7.2. ■

In GMM, instruments z_t are usually seeked out in order to obtain desirable properties, but there is no unique way to search for good instruments, and any choice will expand the dimension of h_t. From Theorem 3.2 (g), this process is bypassed entirely in the MEF. The estimator is unaltered by choice of z_t. To understand this correctly, note that if z_t is in the information set at $t-1$, it is available to the MEF, which then will involve z_t if and only if it is required for optimality.

From Theorem 3.2 (e), MEF is generically more efficient than GMM, based on the same information set and regularity conditions. Indeed, the two cases in Theorem 3.2 (f) where the two estimators coincide are both very special. In many cases, researchers will be interested in a certain set of conditional moment restrictions based on m_t and a certain parameter vector ϕ, but there is no particular reason that the two should be of the same dimension. Even if they were, the condition in (39) is obviously highly special and would probably not be satisfied in any practical application. Further, when the number of parameters and conditional moment restrictions differ, it is at least as likely that the latter is greatest, by Theorem 3.2 again making MEF strictly more efficient than GMM. Finally, if the number of parameters happens to be highest, in many cases it would nevertheless not equal an integer times the number of moment conditions. Even if it did, there is no reason that the left-hand side matrix in (40) should take the sparse form indicated. Certainly, none of these very specialized conditions is satisfied in our application to the AK-Vasicek model, where we calculate $\psi_t^\top (\Psi_t)^{-1}$ explicitly.

Although the forms of the optimal MEF weights (29) and the resulting optimal variance V_{MEF} in (36) are derived here from the theory of estimating functions, based on Godambe and Heyde (1987), and used to improve efficiency of MEF based on m_t relative to GMM based on $h_t = z_t \otimes m_t$, they also appear in the GMM literature on efficiency bounds and optimal instruments that has followed Hansen (1982). Thus, efficiency bounds are derived by Hansen (1985, Lemma 4.3) and Hansen, Heaton, and Ogaki (1988, Theorem 4.2) for general cases where $M_T = \sum_{t=1}^{T} w_t m_t$ is zero-mean but not necessarily a martingale, and the weights (or instruments) w_t are $\dim \phi \times \dim m$ matrices, as in our case. In the martingale case, their bounds coincide with our V_{MEF}. However, these authors do not consider the case that w_t may depend on ϕ, and when seeking estimators achieving the efficiency bound in the restricted class, they further assume conditional homoskedasticity, i.e., the conditional variance is constant across time, $\Psi_t \equiv \Psi$, and ψ_t at the true parameter is known. The latter is no problem for deriving the bound, which is the main focus of these papers, but for a feasible efficient estimator, an initial consistent estimate must be plugged in. The conditional homoskedasticity assumption is restrictive, and we relax this, i.e., our V_{MEF} is a semiparametric efficiency bound in a wider class of models (including conditional heteroskedasticity).

The relation between the results from the efficiency bound and optimal instrument literature on the one hand and standard GMM the way we have defined it on the other hand is not completely direct. We follow Hansen (1982) and consider GMM as the procedure of possibly expanding the moments from m_t to h_t using vector-valued instruments z_t, then minimizing the relevant norm, rather than looking at time-varying matrix-valued instruments w_t.[26] We believe this is in accordance with general practice among applied researchers in macroeconomics and finance. Our results then show that this standard form of GMM is inefficient relative to MEF, which always reaches the general efficiency bound (36). Other estimators reaching the corresponding bound in restricted cases (e.g., imposing conditional homoskedasticity, $\Psi_t \equiv \Psi$) are the other type of GMM estimators, considered in the efficiency bound and optimal instrument literature, with matrix instruments. Of course, by Theorem 3.2 (e), in such cases these estimators coincide with optimal MEF if the restrictions are valid, otherwise MEF is again strictly more efficient.[27]

Part of the reason for the discrepancy between the literatures on estimating equations and optimal estimators on the one hand, and the literature on GMM efficiency bounds and optimal instruments on the other, is that the latter literature seeks semiparametric bounds, for cases where only (conditional) moment conditions are assumed, while the rest of the model is left unspecified (or nonparametrically specified). In such cases, there may not be sufficient information to calculate ψ_t and Ψ_t for the concrete construction of an efficient estimator, although these objects (or special cases of these, like $\Psi_t = \Psi$) enter the received expression for the theoretical bound. In contrast, we have a fully specified model, so we may calculate ψ_t and Ψ_t explicitly, and thus construct the optimal MEF estimator. Indeed, Ψ_t

[26]The analysis in Hansen (1982) does involve a $\dim\phi \times \dim m$ matrix, a_N in Hansen's notation (N corresponds to our T), but this is constant through time and corresponds to G in our discussion before Definition 3.1, see (24). Another paper in this (optimal GMM) tradition is Bates and White (1993).

[27]The form of the weights (29) also resembles that in Newey (1990), but he considers the i.i.d. cross-section case, where the conditioning in ψ_t is on i.i.d. regressors, not the past of a time series as in our case, and he assumes conditional homoskedasticity, $\Psi_t \equiv \Psi$, too. Robinson (1991) allows some restricted forms of serial dependence, but similarly imposes conditional homoskedasticity. Kuersteiner (2001) considers related estimators for linear time series (ARMA) models. In the cross-section case, Chamberlain (1987) uses optimal GMM on an expanding set of moment conditions to approach the efficiency bound. By applying Theorem 3.2 to the special case of a cross-section (including regressors in the information set), our results generalize those available for this case to cover conditional (on regressors) heteroskedasticity, and efficiency is achived by a given estimator (optimal MEF), not just approached in the limit by a sequence of estimators.

is time-varying in the DSGE model, so this is a case of conditional heteroskedasticity, and our estimator is strictly more efficient than any from the literature. We now proceed to this construction.

3.2 MEF with three conditional moment restrictions

We start from the martingale difference sequence in (21), where we approximate the integrals by summation over days between $t-\Delta$ and t. This allows computing $m_t(\phi)$ at trial values of the structural parameters $\phi = (\kappa, \gamma, \eta, \rho, \delta, \sigma)^\top$. To construct the MEF (27), we need the weights w_t in (29), which depend on the conditional mean of the parameter derivatives, ψ_t, and the conditional variance, Ψ_t, of m_t. We have the conditional variances $\Psi_{t,11} = \sigma^2\Delta$, $\Psi_{t,22} = \eta^2 E_{t-\Delta}(\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)^2 dv) + \sigma^2\Delta$, and $\Psi_{t,33} = \eta^2(1-e^{-2\kappa\Delta})/(2\kappa)$. Similarly, the conditional covariances are $\Psi_{t,12} = \sigma^2\Delta$, $\Psi_{t,13} = 0$, and $\Psi_{t,23} = \eta^2 e^{-\kappa\Delta} E_{t-\Delta}\left((\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2) dB_v)(\int_{t-\Delta}^t e^{\kappa(v-(t-\Delta))} dB_v)\right)$. Since analytical expressions are not available, we use Euler approximations for $\Psi_{t,22}$ and $\Psi_{t,23}$,

$$\Psi_t = \begin{pmatrix} \sigma^2\Delta & \sigma^2\Delta & 0 \\ \sigma^2\Delta & \sigma^2\Delta + \eta^2\Delta/(r_{t-\Delta}^f + \delta + \sigma^2)^2 & \eta^2 e^{-\kappa\Delta}\Delta/(r_{t-\Delta}^f + \delta + \sigma^2) \\ 0 & \eta^2 e^{-\kappa\Delta}\Delta/(r_{t-\Delta}^f + \delta + \sigma^2) & \frac{1}{2}\eta^2(1-e^{-2\kappa\Delta})/\kappa \end{pmatrix}. \quad (41)$$

Evidently, Ψ_t depends on $r_{t-\Delta}^f$ and so is time-varying, i.e., this is a case of conditional heteroskedasticity, and so the MEF estimator we construct is strictly more efficient than estimators based on conditional homoskedasticity as in the existing efficiency bound and optimal instrument literature. Note that consistency and asymptotic variance are unaffected by our approximations because these only affect the weights (29).[28] For the martingale increments (21), the parameter derivatives $(\partial m_t/\partial\phi^\top)^\top$ are given by

$$\begin{pmatrix} 0 & \Delta - \gamma\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)dv & -\Delta e^{-\kappa\Delta}\gamma + \Delta e^{-\kappa\Delta} r_{t-\Delta}^f \\ 0 & -\kappa\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)dv & -(1-e^{-\kappa\Delta}) \\ 0 & \eta\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)^2 dv & 0 \\ \Delta & \Delta & 0 \\ 0 & \kappa\gamma\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)^2 dv - \eta^2\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)^3 dv & (1-e^{-\kappa\Delta}) \\ -\sigma\Delta & \begin{array}{c} -\sigma\Delta + 2\sigma\kappa\gamma\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)^2 dv \\ -2\sigma\eta^2\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)^3 dv \end{array} & 2\sigma(1-e^{-\kappa\Delta}) \end{pmatrix}. \quad (42)$$

Now apply conditional expectation to get $\psi_t = E_{t-\Delta}(\partial m_t/\partial\phi^\top)$, interchange the order of integration in (42), and use the deterministic Taylor expansion (e.g., Aït-Sahalia, 2008),

[28] Unaffected asymptotic variance requires that the effect of the Euler approximation wears off asymptotically, otherwise the expression may be adjusted accordingly.

which reads

$$E(g(r_s)|r_u) = \sum_{i=0}^{k} \frac{\Delta^i}{i!} A^i g(r_u) + O(\Delta^{k+1}), \quad s \geq u \tag{43}$$

where A is the infinitesimal generator in the Vasicek model, $Ag(x) = \kappa(\gamma - x)g'(x) + \frac{1}{2}\eta^2 g''(x)$. The function $g(\cdot)$, for example $g(x) = 1/x$ in $\psi_{t,21}$, must be sufficiently smooth. For example, a first-order Taylor expansion, $k = 1$, yields the approximation[29]

$$\begin{aligned}
\int_{t-\Delta}^{t} E_{t-\Delta}(1/r_v)dv &= \int_{t-\Delta}^{t} E_{t-\Delta}(1/r_v)dv \\
&\approx \int_{t-\Delta}^{t} \left(1/r_{t-\Delta} + (v - (t-\Delta))\left(-\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^2 + \eta^2/r_{t-\Delta}^3\right)\right) dv \\
&= \Delta/r_{t-\Delta} - (t-\Delta)\left(-\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^2 + \eta^2/r_{t-\Delta}^3\right)\Delta \\
&\quad + \tfrac{1}{2}(t^2 - (t-\Delta)^2)\left(-\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^2 + \eta^2/r_{t-\Delta}^3\right) \\
&= \Delta/r_{t-\Delta} - \left(\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^2 - \eta^2/r_{t-\Delta}^3\right)\tfrac{1}{2}\Delta^2,
\end{aligned}$$

in which $r_v = r_v^f + \delta + \sigma^2$. Thus expanding all terms involving integrals in (42), we have the transpose of the conditional mean parameter derivative $\psi_t^\top$ given explicitly as

$$\begin{pmatrix}
0 & \Delta - \gamma\left(\Delta/r_{t-\Delta} - \left(\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^2 - \eta^2/r_{t-\Delta}^3\right)\frac{1}{2}\Delta^2\right) & \Delta e^{-\kappa\Delta}(r_{t-\Delta}^f - \gamma) \\
0 & -\kappa\left(\Delta/r_{t-\Delta} - \left(\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^2 - \eta^2/r_{t-\Delta}^3\right)\frac{1}{2}\Delta^2\right) & -(1 - e^{-\kappa\Delta}) \\
0 & \eta\left(\Delta/r_{t-\Delta}^2 - \left(2\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^3 - 3\eta^2/r_{t-\Delta}^4\right)\frac{1}{2}\Delta^2\right) & 0 \\
\Delta & \Delta & 0 \\
0 & \begin{array}{c}\kappa\gamma\left(\Delta/r_{t-\Delta}^2 - \left(2\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^3 - 3\eta^2/r_{t-\Delta}^4\right)\frac{1}{2}\Delta^2\right) \\ -\eta^2\left(\Delta/r_{t-\Delta}^3 - \left(3\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^4 - 6\eta^2/r_{t-\Delta}^5\right)\frac{1}{2}\Delta^2\right)\end{array} & (1 - e^{-\kappa\Delta}) \\
-\sigma\Delta & \begin{array}{c}-\sigma\Delta + 2\sigma\kappa\gamma\left(\Delta/r_{t-\Delta}^2 - \left(2\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^3 - 3\eta^2/r_{t-\Delta}^4\right)\frac{1}{2}\Delta^2\right) \\ -2\sigma\eta^2\left(\Delta/r_{t-\Delta}^3 - \left(3\kappa(\gamma - r_{t-\Delta})/r_{t-\Delta}^4 - 6\eta^2/r_{t-\Delta}^5\right)\frac{1}{2}\Delta^2\right)\end{array} & 2\sigma(1 - e^{-\kappa\Delta})
\end{pmatrix}, \tag{44}$$

in which, again, $r_{t-\Delta} = r_{t-\Delta}^f + \delta + \sigma^2$. This completes the construction of the optimal martingale estimating function $M_T = \sum_t \psi_t^\top(\Psi_t)^{-1} m_t$. The condition $M_T(\phi) = 0$ involves the same number of equations and unknowns, and is solved exactly for the optimal MEF estimator $\widehat{\phi}$. The asymptotic distribution is given by (35)-(36). Again, this is a case of a strict efficiency gain, relative to estimators from the literature.

3.3 MEF extensions

So far, we have considered the case where all variables in the system are observable, albeit using some mixed-frequency properties of the data. The MEF approach can be generalized

[29] A simpler Euler approximation would neglect all second order terms. The Taylor expansion shown improves identification as more structural parameters appear. The relevance of these terms, however, will be model-specific. For our baseline model the empirical results based on $\Delta^2 = 0$ are similar.

to the empirically relevant cases of latent variables (e.g., unobserved real interest rates, stochastic discount rates, stochastic volatility, regime-switching models, etc.) and of mixed-frequency estimation with a different frequency for each series. To illustrate, we consider four extensions, set in the context of the AK-Vasicek model: (i) The daily interest rate r_t is latent and cannot be backed out from data – there is no daily observed series proxying for this variable; (ii) the interest rate r_t is subject to regimes with high and low volatility; (iii) The interest rate r_t is subject to stochastic volatility; (iv) output Y_t is observed at a lower (say, quarterly) frequency than consumption C_t (say, monthly).

Cases (i) to (iii) serve to illustrate our approach to latent variables (the interest rates, the volatility regime, respectively the stochastic volatility are latent). Cases (i) and (iv) illustrate our handling of truly missing data (interest rates, respectively parts of the output series). Case (i) is motivated by the concern that expected inflation and hence the real rate of interest should be treated as missing in some applications. An interesting feature of the approach is that we may infer the latent series and conduct a model specification check. Cases (ii) and (iii) show how our approach can be applied in cases when the financial data display patterns of stochastic volatility and/or regime switching. Case (iv) reinforces our use of data sampled at mixed frequencies. Output may be proxied by industrial production at the monthly frequency (see below), but it may be of interest to compare with results using actual output, available only quarterly. In the latter case, consumption need not be aggregated to quarterly frequency.

3.3.1 Latent interest rate

For the latent interest rate generalization, case (i), note that in the MEF approach with complete data the condition $E\left(M_T\right)=0$ is satisfied at the true parameter value, where $M_T=\sum_{t=1}^{T} w_t m_t$. In the incomplete data setting, define $\mathcal{F}_t$ as the information set generated by $\{C_s, Y_s\}_{s=1}^{t}$ (but not the missing interest rates). By $E\left(M_T\right)=0$ and iterated expectations, we have $E\left(\sum_t w_t E\left(m_t|\mathcal{F}_{t-1}\right)\right)=0$, as the weights w_t depend only on information through $t-1$. Thus, in the estimation, we may replace the moments m_t by their conditional expectations given $\mathcal{F}_{t-1}$. For example, for our application to the stochastic AK-Vasicek model, we replace the integrals involving the daily interest rate by conditional expectations given monthly or quarterly interest rate proxies, based on the information set. In theory, this involves an efficiency loss, since the resulting procedure is formally a variant of standard GMM, i.e., the basis of the approach is an unconditional zero mean condition, not a conditional one. However, we show in the sequel that the approach works surprisingly well in cases we consider.

The new moments for estimation, say, $m_t^* = E(m_t|\mathcal{F}_{t-\Delta})$, are given by

$$\begin{pmatrix} \ln(C_t/C_{t-\Delta}) - E\left(\int_{t-\Delta}^t r_v dv | r_{t-\Delta}^*\right) + \left(\rho + \delta + \frac{1}{2}\sigma^2\right)\Delta \\ \ln(Y_t/Y_{t-\Delta}) + \left(\kappa + \rho + \frac{1}{2}\sigma^2 + \delta\right)\Delta \\ -E\left(\int_{t-\Delta}^t r_v dv + \kappa\gamma \int_{t-\Delta}^t 1/r_v dv - \frac{1}{2}\eta^2 \int_{t-\Delta}^t 1/r_v^2 dv | r_{t-\Delta}^*\right) \\ r_t^* - (1 - e^{-\kappa\Delta})\gamma - e^{-\kappa\Delta} r_{t-\Delta}^* \end{pmatrix}, \tag{45}$$

where $r_{t-\Delta}^*$ is an interest rate proxy based on consumption and income data through $t - \Delta$. Here, $\Delta = 1/4$ is used in the empirical work. From earlier, the model implies $K_t = Y_t/r_t$ and $C_t = \rho K_t$, so a model-consistent proxy at the macro frequency is $r_t^* = \rho Y_t/C_t$, in which Y_t/C_t is approximated by the ratio of quarterly observed income and consumption data.[30]

One possibility for implementation of the conditional expectations of the integrals in (45), i.e., integrating out the latent interest rate process r_v from the terms $E\left(\cdot|r_{t-\Delta}^*\right)$, is simulation. We refer to the resulting procedure as Simulated MEF, or SMEF. The SMEF approach applies generally to models involving latent variables. Thus, each integral involves drawing a path for r_v from $dr_v = \kappa(\gamma - r_v)dv + \eta dB_v$ using an Euler scheme from $v = t - \Delta$ to t, starting at the proxy value $r_{t-\Delta}^*$, and the expectation is formed by averaging over paths. The interest rate (or latent state variable) is similarly integrated out of $w_t = \psi_t^\top (\Psi_t)^{-1}$, or, in the specific case, $r_{t-\Delta}$ is simply replaced by its proxy $r_{t-\Delta}^*$ in the expressions (44) and (41) for $\psi_t^\top$ and Ψ_t. In the iterative solution of the estimating equation $\sum_t \psi_t^\top (\Psi_t)^{-1} m_t^* = 0$, the parameter dependence (in our model, through ρ) of the implied state variables is accounted for.

3.3.2 Regime-switching

In this section, we incorporate a regime-switching spot rate volatility process into our baseline AK-Vasicek specification, case (ii). In particular, we specify $\mu(r_t) = \kappa(\gamma - r_t)$ and $\eta(r_t) = \eta_t$, where $\kappa > 0$ is the speed and γ the target rate of mean reversion, and η_t a continuous-time Markov process with state space $\Theta \equiv \{\eta_h, \eta_l\}$, where $\eta_h > \eta_l$, and

$$d\eta_t = (\eta_l - \eta_h)dq_{1,t} + (\eta_h - \eta_l)dq_{2,t}.$$

The Poisson process $q_{1,t}$ counts how often the process switches from the high volatility level η_h to the low level η_l, and $q_{2,t}$ is a Poisson processes counting the switches from the low to the high regime, only one process being active at a time. We use state-dependent arrival rates $\phi_1(\eta_t) = \phi_{hl}$ for periods where $\eta_t = \eta_h$ and $\phi_1(\eta_t) = 0$ otherwise, $\phi_2(\eta_t) = \phi_{lh}$ for $\eta_t = \eta_l$ and $\phi_2(\eta_t) = 0$ otherwise. This specification is inspired by financial time series

[30] To retain information, the smoothing in (45) is applied only to terms involving the latent interest rate, and r_t in the third moment is smoothed using $E(\cdot|\mathcal{F}_t)$ rather than $E(\cdot|\mathcal{F}_{t-\Delta})$.

exhibiting periods of high and low volatility. Liu, Waggoner, and Zha (2011) consider a related, discrete-time, regime-switching DSGE model and introduce regime-switches in the standard deviation of the interest rate (among other shocks).[31] One challenge in discrete-time models is the solution of the regime-switching model (cf. Foerster, Rubio-Ramírez, Waggoner, and Zha, 2015), while in the present approach the economic model can be solved analytically.

Similarly to the case of latent interest rates, our approach for the regime-switching model is to derive moments for estimation, say, $m_t^* = E\left(m_t|\mathcal{F}_{t-\Delta}\right)$, given by (see web appendix Section C.1)

$$\begin{pmatrix} \ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^t r_v^f dv + \left(\rho - \frac{1}{2}\sigma^2\right)\Delta \\ \ln(Y_t/Y_{t-\Delta}) - \int_{t-\Delta}^t r_v^f dv + \left(\kappa + \rho - \frac{1}{2}\sigma^2\right)\Delta - \kappa\gamma \int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)dv \\ +E\left(\frac{1}{2}\int_{t-\Delta}^t \eta_v^2/(r_v^f + \delta + \sigma^2)^2 dv|\mathcal{F}_{t-\Delta}\right) \\ r_t^f - (1-e^{-\kappa\Delta})(\gamma - \delta - \sigma^2) - e^{-\kappa\Delta} r_{t-\Delta}^f \end{pmatrix}. \tag{46}$$

One possibility for implementation of the conditional expectation of the integral in (46), i.e., integrating out the latent volatility process η_v from the term $E\left(\frac{1}{2}\int_{t-\Delta}^t \eta_v^2/(r_v^f + \delta + \sigma^2)^2 dv|\mathcal{F}_{t-\Delta}\right)$, is simulation. This is the idea of the SMEF approach. For an off-the-shelf implementation, because this is not the main focus of the paper, we follow an alternative route and use an Euler approximation of the integral, complementing the MEF approach with a filtering method to filter the latent volatility state from the data at time t. The Euler approximation of the integral is

$$E\left(\frac{1}{2}\int_{t-\Delta}^t \eta_v^2/(r_v^f + \delta + \sigma^2)^2 dv|\mathcal{F}_{t-\Delta}\right) \approx \frac{1}{2}\Delta(\eta_{t-\Delta}^*)^2/(r_{t-\Delta}^f + \delta + \sigma^2)^2, \tag{47}$$

where $\eta_{t-\Delta}^* = E(\eta_{t-\Delta}|\mathcal{F}_{t-\Delta})$ is the filtered volatility state based on interest rate data through $t-\Delta$. We use the Hamilton (1989) filter to draw inference on the latent regime, thus producing the probability that the latent volatility process is in a given state, say, the high, at any given point in time, $p(\eta_t = \eta_h|\mathcal{F}_{t-\Delta})$. This probability is used to approximate the expectation in (47), and therefore (46), either by taking the high-volatility state, $\eta_{t-\Delta}^* = \eta_h$, whenever the probability $p(\eta_t = \eta_h|\mathcal{F}_{t-\Delta}) > 0.5$ (used here), or using the probability to calculate an expected value in between of η_l and η_h.[32]

[31] Liu, Waggoner, and Zha (2011) implement the Sims, Waggoner, and Zha (2008) algorithm on the log-linearized equilibrium conditions, and find Bayesian estimates (modes of the posterior distributions) using quarterly data for the high and low volatility regimes of 0.004 and 0.001, respectively.

[32] We use the likelihood of the corresponding discrete-time Ornstein-Uhlenbeck process for the interest rate in the Hamilton filter. This filter is run alongside the MEF approach, and provides the filtered interest rate volatility states depending, among others, on the regime-switching parameters η_h, η_l, ϕ_{lh} and ϕ_{hl}. In web appendix Section A.3 we derive the transition probability matrix for the continuous-time Markov chain of the regime-switching model and, conversely, show how to back out the instantaneous transition rates of the Poisson process from any given probability matrix.

3.3.3 Stochastic volatility

Following Andersen and Lund (1997) we incorporate stochastic volatility into our baseline AK-Vasicek specification, case (iii). In particular, we specify $\mu(r_t) = \kappa_1(\gamma_1 - r_t)$ and $\eta(r_t) = \eta_t$, where $\kappa_1 > 0$ is the speed and γ_1 the target rate of mean reversion, and η_t the time-varying stochastic volatility (SV) process following

$$d\log(\eta_t^2) = \kappa_2(\gamma_2 - \log(\eta_t^2))dt + \xi dW_t.$$

Here, $\kappa_2 > 0$ is the speed and γ_2 the target rate of mean reversion in (log-)volatility, the value of ξ governs its variability (the volatility-of-volatility term), and W_t is an independent Brownian motion process. Some evidence of stochastic volatility in (discrete-time) DSGE models has been documented by Fernández-Villaverde, Guerrón-Quintana, and Rubio-Ramírez (2015). Below, we show how the continuous-time approach and the MEF method can deal with the estimation of such models and how high-frequency data help proxying the latent volatility process.

Similarly to the prior extensions of the baseline MEF method, our approach is to derive moments for estimation, say, $m_t^* = E\left(m_t|\mathcal{F}_{t-\Delta}\right)$, given by (see web appendix Section C.2)

$$\begin{pmatrix} \ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^t r_v^f dv + \left(\rho - \frac{1}{2}\sigma^2\right)\Delta \\ \ln(Y_t/Y_{t-\Delta}) - \int_{t-\Delta}^t r_v^f dv + \left(\kappa_1 + \rho - \frac{1}{2}\sigma^2\right)\Delta - \kappa_1\gamma_1 \int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)dv \\ + \frac{1}{2}E\left(\int_{t-\Delta}^t \eta_v^2/(r_v^f + \delta + \sigma^2)^2 dv|\mathcal{F}_{t-\Delta}\right) \\ r_t^f - (1 - e^{-\kappa_1\Delta})(\gamma_1 - \delta - \sigma^2) - e^{-\kappa_1\Delta} r_{t-\Delta}^f \\ 2\log(\eta_t^*) - (1 - e^{-\kappa_2\Delta})\gamma_2 - e^{-\kappa_2\Delta} 2\log(\eta_{t-\Delta}^*) \end{pmatrix}, \quad (48)$$

where $\eta_{t-\Delta}^*$ is an interest rate volatility proxy based on interest rate data through $t - \Delta$. Thus, we now use four moments, instead of three. One possibility for implementation of the conditional expectation of the integral in (48) is simulation, which is the idea of the SMEF approach. For ease of implementation, we use an Euler approximation of the expectation $E\left(\cdot|\mathcal{F}_{t-\Delta}\right)$ in $E\left(\int_{t-\Delta}^t \eta_v^2/(r_v^f + \delta + \sigma^2)^2 dv|\mathcal{F}_{t-\Delta}\right)$ and evaluate the moments by using the volatility proxy series. We use the realized volatility estimate based on daily interest rate data, at either the monthly ($\Delta = 1/12$) or quarterly frequency ($\Delta = 1/4$), such that

$$\eta_t^* = \sqrt{\frac{1}{\Delta}\sum_{i=1}^{P-1}(r_{t-\Delta+(i+1)\Delta/P}^f - r_{t-\Delta+i\Delta/P}^f)^2},$$

where P is the number of days in the period as in Section 2.5. The high-frequency availability of financial data thus provides the proxy at the relevant monthly and quarterly frequency, with which (48) can be estimated at either frequency. The use of daily data to assess volatility over longer intervals dates back to Merton (1980) and French, Schwert, and Stambaugh

(1987). Our use of the realized volatility approach is inspired by Bollerslev and Zhou (2002).[33] We use the volatility proxy η_t^* in both (48) and the corresponding ψ_t and Ψ_t matrices.

3.3.4 Mixed frequency

The mixed frequency extension, case (iv), where consumption is available on a monthly basis and output only quarterly is slightly different. Here, a complete (monthly) output proxy series Y_t^* is simply constructed recursively by letting $Y_t^* = Y_t$ in the (quarterly) periods where output data are available, and

$$Y_t^* = \exp\left(\ln(Y_{t-\Delta}^*) + \int_{t-\Delta}^t r_v dv - \left(\kappa + \rho + \delta + \tfrac{1}{2}\sigma^2\right)\Delta + \kappa\gamma \int_{t-\Delta}^t 1/r_v dv - \tfrac{1}{2}\eta^2 \int_{t-\Delta}^t 1/r_v^2 dv\right)$$

in the intra-quarter periods when output is missing. Here, $r_v = r_v^f + \delta + \sigma^2$, using observed daily r_v^f. This is model consistent prediction, not simulation. In particular, the resulting proxy series Y_t^* depends on the parameters. The proxy series is now substituted for Y_t in the original estimating equation $\sum_t w_t m_t = 0$. We refer to this procedure as mixed-frequency MEF, or MF-MEF. When solving for the parameter estimates, the dependence of the constructed output proxy series on trial parameter values is again accounted for.

Both generalized approaches, SMEF and MF-MEF, are akin to filtering. Thus, in the presence of latent variables, cases (i) through (iii), SMEF recasts m_t in the estimating equation in terms of a set of conditional expectations or filtered predictions, given the information actually available (Euler-approximations may bypass simulations). For mixed-frequency estimation, case (iv), MF-MEF replaces missing data by conditional predictions given the actual observations. In both cases, standard errors may be calculated using the bootstrap.

4 Simulation Study

To assess the estimation methods from the previous section we run a simulation experiment. We first detail the set-up of our analysis. As in the previous section, our illustration is based on the AK-Vasicek model. In the text below we focus on the case with observable variables, using three conditional moment restrictions for estimation. In Section D of the web appendix we provide additional simulation evidence with discussion of and results for MEF extensions to the cases of latent variables and mixed frequency data, time invariance, the benefits of high-frequency data, and the comparison to discrete time. Several additional appendix tables

[33] The estimation of volatility from high-frequency financial data has attracted immense attention, see, e.g., Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002), and Zhang, Mykland, and Aït-Sahalia (2005). As sampling frequency increases, realized volatility under wide conditions converges to integrated volatility, $(\eta_t^*)^2 \to \int_{t-\Delta}^t \eta_v^2 dv$ as $P \to \infty$.

that we refer to below, including results using five conditional moment restrictions (based on both first and second order moments of the model residuals), are available in Section E of the web appendix.

4.1 Set-up

We simulate 25 years of both monthly and quarterly data from the model. We use simple Euler approximations to the differential equations in (13). The step length of the Brownian terms is taken as 1/3000. This corresponds to dividing each of the 12 months of the year into 25 days, each in turn consisting of 10 periods.

There are two further computational issues when simulating from the model: Obtaining the integrals involving the interest rate and initialization of the simulations. Concerning the first issue, we obtain the monthly integrals over the interest rate, denoted with $\int_{t-\Delta}^{t} g(r_v)dv$ where $\Delta = 1/12$ (see Section 2.5), by taking the average of the functions $g(r_v)$ over the 25 simulated days per month. For example, $\int_{t-\Delta}^{t} 1/(r_v^f + \delta + \sigma^2)dv$ for the monthly simulations is approximated by $(\sum_{i=1}^{25} 1/(r_{t-\Delta+i\Delta/25} + \delta + \sigma^2))\Delta/25$. For the quarterly simulated data we use a similar approximation, but now over the 75 days in the Euler approximation. Concerning the second issue, we normalize initial output to unity and initialize the other variables consistently with our model: $\ln Y_0 = 0$, $r_0 = \gamma$, and $\ln C_0 = \ln(\rho Y_0/r_0)$.

We generate 1,000 data sets and estimate the parameters according to the approaches of Section 3. In particular, we report the parameter estimates for the OLS, FGLS-SUR-IV, GMM, and MEF methods. In the first two cases, we use the minimum distance approach to get the structural parameters from the reduced form estimates. We choose DGP parameter values in a way roughly corresponding to empirical estimates obtained in Section 5.2 below. In particular, we use $\kappa = 0.2$, $\gamma = 0.1$, $\eta = 0.01$, $\rho = 0.03$, $\delta = 0.05$, and $\sigma = 0.02$ (see column DGP in Table 1).[34] In the web appendix, we report the sensitivity of our estimation methods to different DGP values (cf. Table E1).

4.2 Simulation results: MEF with three conditional moment restrictions

[insert Table 1]

Table 1 provides the results of the simulation study. In the first column we list the parameter values as they are used in the data generating process (DGP), in columns 2 through 5 the

[34]This corresponds to $\tilde{\kappa} = 0.0165$, $\tilde{\gamma} = 0.0083$, $\tilde{\eta} = 0.0002$, $\tilde{\beta} = 0.9975$, $\tilde{\delta} = 0.0042$, and $\tilde{\sigma} = 0.0058$ for the discrete-time model at monthly frequency with $\Delta = 1/12$, and $\tilde{\kappa} = 0.0488$, $\tilde{\gamma} = 0.025$, $\tilde{\eta} = 0.0012$, $\tilde{\beta} = 0.9925$, $\tilde{\delta} = 0.0124$, and $\tilde{\sigma} = 0.0101$ for the discrete-time model at quarterly frequency with $\Delta = 1/4$.

estimates obtained from the simulated monthly data, and in columns 6-9 the estimates from the quarterly data. For our four estimation methods we provide the median estimate of each parameter, and below the interquartile range of the 1,000 estimates. Not all six structural parameters are identified in the regression-based estimation methods, when not exploiting second moments (in particular, the residual variances of $\varepsilon_{C,t}$ and $\varepsilon_{r,t}$). In fact, five parameter combinations are identified, so one possibility is to fix one parameter at the outset and estimate the remaining five. There is some choice regarding which parameter to set, since the two combinations $\rho - \frac{1}{2}\sigma^2$ and $\delta + \sigma^2$ are identified, i.e., either ρ, δ, or σ^2 could be restricted. We choose to set δ to the DGP value of 0.05, which economically is interpreted as depreciation of physical capital of 5 percent per year, in the regression-based approaches. The same identification issue occurs for the GMM when the standard lagged right-hand side variables are used as instruments and no higher order moments are used. Also here we fix δ at the outset to the known DGP value of 0.05.

Overall, while simple OLS, which ignores estimation problems such as contemporaneous cross-equation correlation of errors and endogeneity of right-hand side variables, has some trouble identifying the structural parameters, the FGLS-SUR-IV, GMM, and MEF approaches produce estimates of γ, η and ρ that are remarkably close to the values in the DGP. Without exploiting further moments, σ seems to be only weakly identified using GMM, as reflected by the large inter-quartile range, and generates heavily biased point estimates in the regression-based approaches. In the MEF approach, σ is already identified from considering only three moment conditions. Here, the identification of all structural parameters, including δ, works through the optimally chosen weight (or instrument) matrices involving the conditional mean parameter derivatives and conditional variances of the martingale increments. Similar results hold for both monthly and quarterly data.

The mean-reversion parameter κ of the Vasicek specification is hard to estimate. Here, a value 0.2 is used in the DGP, while the median estimates are in the range from 0.30 to 0.35 using monthly data, and only slightly better in quarterly data. This upward bias in the mean-reversion parameter estimate is well established (see Tang and Chen, 2009; Wang, Phillips, and Yu, 2011). In particular, for values of κ close to zero, i.e., a near unit root situation typical of many financial time series (Yu, 2012), a bias correction may be preferable.[35] We find a similar bias across different estimation approaches, and for different sets of DGP

[35] We apply bias correction methods from Tang and Chen (2009) and Yu (2012). Although formally the methods are not directly applicable, they perform reasonably well, providing monthly and quarterly bias-corrected MEF median estimates of κ at values 0.192 and 0.273, respectively. The bootstrapped bias correction along the lines of Tang and Chen (2009, Section 4), which we have adjusted to our setting, yields a bias-corrected MEF median estimate of κ of 0.204 (cf. Table E2) and will be used for the empirical application.

values (cf. Table E1). This finite-sample bias, however, does not seem to translate to other structural parameters (cf. Figure 1). Moreover, given the relatively wide inter-quartile range, the κ estimates are still within reasonable distance.

[insert Figure 1]

In Figure 1 we provide the histograms of the 1,000 estimates that we obtain for the parameters using the MEF approach on both monthly data (Panel A) and quarterly data (Panel B). The figure confirms the findings from Table 1: The parameters γ, η, ρ, δ, and σ are centered close to the DGP values. In addition, it becomes clear that the modes of the histograms for κ in fact are quite close to the DGP values, but the distributions are skewed, thus causing the difference between median estimates and DGP values reported in Table 1.

We also implement minimum distance (cf. Section A.1.4), using the residual variance from the consumption (not reported), or both the consumption and interest rate equation as additional moments (cf. Table E3 in the web appendix), along with β in the regression-based approaches, allowing better identification of σ for monthly data (columns 2 and 3) respectively quarterly data (columns 6 and 7). A similar idea for the GMM and MEF methods is to include further moment restrictions in the martingale estimating equation (27). The results based on five moment restrictions (instead of three), with the conditional moments derived in web appendix Section C.3, are reported in Table E3, columns 4 and 5, for monthly data, respectively 8 and 9 for quarterly data. Including more moments indeed yields (better) identification of the six parameters. The relatively poor performance of the regression-based approaches clearly relates to the restriction to first moments. In cases where the econometrician is able to use second moments, it is advisable to so.

Taken together, the simulation study indicates that the GMM and MEF approaches are successful in recovering parameter estimates from the data. The regression-based methods exhibit reasonable performance, but only after accounting for potential estimation problems due to cross-equation correlation and endogeneity. In empirical work, the regression-based approaches would also require iteration over the proxy $\hat{r}_t$ in (17). In our simulation study, the values for δ_0 and σ_0 in (17) are set to their corresponding DGP values of δ and σ.[36]

5 Data and Results

In this section we estimate the AK-Vasicek model with logarithmic preferences based on empirical data. We provide results for the estimation approaches developed in Section 3, using US mixed frequency macro and financial data.

[36] We examine the sensitivity of OLS and FGLS-SUR-IV parameter estimates to δ_0 and σ_0 in Table E4.

5.1 Data

[insert Figure 2]

To estimate the system (15) we need data on production, consumption, and the short rate. We obtain these data for the US from the Federal Reserve Economic Dataset (FRED), maintained by the Federal Reserve Bank of St. Louis. To measure production, we use both real Industrial Production (IP), available at the monthly level, and real Gross Domestic Product (GDP), available at the quarterly level. We use real Personal Consumption Expenditures (PCE) at the monthly and the quarterly level to proxy consumption. In Figure 2 we show the plots of the monthly and quarterly growth rates of the variables (Panels A and B). Our data set spans the period from January 1982 to December 2012.

We combine the data on these aggregate macro series with financial data available at higher frequency, in particular, the short rate. This rate is a theoretical concept and corresponds to an infinitesimal time to maturity. In applied work, the short rate is sometimes treated as a latent variable that is filtered from observed yield data (e.g., De Jong, 2000). As a starting point, we follow Chapman, Long, and Pearson (1999), and use the 3-month interest rate as a proxy for the short rate r_t^f of the risk-free financial asset, here taken as the US treasury bonds. This interest rate is available from the FRED data set at daily frequency. We use this series to obtain our monthly and quarterly figure by taking the last observation in the relevant period. Panel (C) of of Figure 2 shows the daily interest rate series. In the series, a general downward trend of the interest rate is evident.

Finally, we use the interest rate series to compute approximations to the integrals that appear in our empirical specification. We approximate the monthly and quarterly series of integrals using the daily spot rate observations. Given system (15), we approximate three integrals: $\int_{t-\Delta}^{t} g(r_v^f)dv \approx \Delta/P \sum_{i=1}^{P} g(r_{t-\Delta+i\Delta/P}^f)$, where $r_{t-\Delta+i\Delta/P}^f$ is the 3-month rate on day i of period t, and P the number of days in the period between $t-\Delta$ and t.

5.2 Estimates: MEF with three conditional moment restrictions

When taking our model to the empirical data, we experienced numerical optimization difficulties with some of the parameters, which may be due to possible model misspecification. From Section 4.2, five parameter combinations are in theory identified without exploiting second moments (specifically, the consumption growth residual variance). Hence, setting one parameter should allow estimating the remaining five, but the regression-based methods and GMM had difficulty doing so. The iterative optimization routines either diverged or produced economically unreasonable estimates that furthermore depended heavily on starting values. Consequently, in the reported results, two of the six parameters are set at pre-fixed values for these methods, instead of just one. Again, since $\rho - \frac{1}{2}\sigma^2$ and $\delta + \sigma^2$ are identified, either two of ρ, δ, or σ^2 could be restricted. From the simulation study, σ is weakly identified when not exploiting second moments, so we set this at 0.02, and δ again at 0.05 (cf. also Table E5 in the accompanying web appendix, where we use variance terms for the regression-based methods and five moments for both GMM and MEF).

In the MEF approach, all six parameters are identified even without using second moments. Unlike in the simulations, we implemented a slightly simplified version of MEF in order to avoid similar problems as those encountered with the other methods. Essentially, the optimal MEF weights $\psi_t^\top(\Psi_t)^{-1}$ from (29) were replaced by weights $\psi_t^\top(\hat{\Psi})^{-1}$, i.e., still with time-varying conditional mean parameter derivatives of martingale increments (31), but the conditional variance (30) replaced by a constant estimate. Similarly to optimal two-step GMM, we first estimated with $\Psi = I_3$, then with $\hat{\Psi}$ computed as the outer product of the fitted residuals at time t from the first step. We label this approach two-step MEF. All six parameters were successfully estimated in both steps, and the values for δ and σ broadly in line with the pre-specified values used in the other methods.[37]

These identification problems can be interpreted as a first indication that the AK-Vasicek specification with logarithmic preferences probably will not match the data well. This leads to some degree of problems for all estimation procedures, in particular GMM, although less so for the MEF. The likely model misspecification is discussed further below in Section 5.3.

[insert Table 2]

Table 2 provides structural parameter estimates based on both monthly and quarterly data, using industrial production respectively GDP for output, obtained using the OLS, FGLS-SUR-IV, GMM, and MEF approaches. The regression-based estimation methods provide fairly similar estimates of the four parameters for monthly and quarterly data. By the point estimates, the short rate is mean-reverting, but not very strongly, with speed parameter κ around 0.08 (0.047 to 0.109 for quarterly data).[38] A speed of zero implies a unit root. The implied first order autocorrelation is $e^{-0.08\times 1/12} = 0.99$ for monthly data. The long-term target rate γ is about 10%, and the volatility η of the short rate innovation is between 1.3% and 2.6%. As is well-known, the interest rate has been declining during the period (see Figure 2, Panel C), so the model will not yield a good fit and is likely misspecified,

[37] In the web appendix we show that the two-step MEF approach yields similar results compared to MEF with optimal weights (cf. Table E6). This two-step approach corresponds to the conditional homoskedasticity assumption from the efficiency bound and optimal instrument literature, but we have shown that our model is conditionally heteroskedastic, and optimal MEF therefore theoretically most efficient, which could show up in other applications, in other data, or more elaborate conditionally heteroskedastic models.

[38] The bootstrapped bias-corrected MEF median estimate of κ is at 0.084 (0.109 for quarterly data).

but it is worth noting that the 10% level makes sense. Thus, by the asset pricing equation (14), it comprises the average risk-free rate from the data, the risk premium σ^2 consistent with logarithmic preferences, and the rate δ of physical capital depreciation. For the given δ and σ, the time preference parameter ρ is estimated at around 1% in monthly and 2% in quarterly data. Of course, it is important to note that the two data sets differ not only by sampling frequency, but also by relying on industrial production respectively GDP.

The GMM estimates of κ and γ are similar to those from the regression-based methods, whereas the point estimates of ρ are about 0.5 percentage points smaller. The main difference is that GMM does not pick up any of the innovation variance in the interest rate process. The MEF results are slightly different in some respects, and it should be kept in mind that they are obtained without restricting δ and σ. In particular, in quarterly data, the depreciation rate δ is estimated at 6.2%, i.e., 25% higher than the pre-set value used for the other methods. Consistently with this, the long-run mean interest rate γ is higher, too, at 13%. In monthly data, these are lower, at 2.5% respectively 5.1%, and the κ and η estimates at both frequencies are similar to those from other methods (except that GMM had trouble estimating η). Further, MEF produces precise estimates of ρ at the quarterly frequency and σ at the monthly, even significant at conventional levels (many of the received estimates are statistically insignificant, across all parameters and methods). Indeed, both the monthly and the quarterly MEF estimate of σ confirms the value 0.02 imposed in the other methods.

5.3 Estimates: MEF extensions

[insert Table 3 and Figure 3]

Table 3 shows the results for our MEF extensions: the regime-switching and the stochastic volatility models (columns 2 and 3), and the latent short rate and mixed-frequency estimation (columns 5 and 6). For comparison, we replicate in columns 1 and 4 the MEF results from Table 2. The latent variable extension, case (i), reveals more evidence on the sources of misspecification: The counter-factual model-implied short rate when it is compared to the observed interest rate proxy. The results of SMEF in column 5 show that if parts of the data were model-generated, the interest rate consistent with macro dynamics would show much higher mean reversion κ, about 13%, and a much lower long-term target rate γ, at about 5.8%, with very small innovation variance, η close to zero. The high t-values reflect the fact the standard errors are likely downward biased (so the t-values upward biased), because the additional uncertainty from drawing the short rate is not taken into account, which in principle could be accounted for by bootstrapping. For illustration, in Panel A of Figure 3 we plot one simulated short rate path r_t^*. It is worth noting that in contrast to the observed risk-free rate r_t^f, any model-implied short rate r_t^*, for the given consumption and income data, will be upward sloping. Because the proxy r_t^* is set to the ratio $\rho Y_t/C_t$ at quarterly frequency (indicated by dots), the model-implied short rate follows the same pattern. For this reason, SMEF is not applicable to monthly empirical data since IP is measured as an index.

The latent variable extension with regime switches of the short rate volatility, case (ii), indicates that there is not much mean reversion in the interest rate, κ about 0.6%, but rather switches between high and low volatility states. From the Hamilton filter we obtain a low volatility regime, η_l, at about 0.5% and a high volatility regime, η_h, at about 2.1%. The instantaneous transition rates ϕ_{hl} and ϕ_{lh} imply an annual transition probability matrix with a probability of staying in the low regime next year (based on monthly estimates) of 74.9% (given a low volatility regime today) and with a probability of staying in the high regime next year of about 37.9% (given a high volatility regime today). Panel B of Figure 3 shows that the recent financial crisis starting in 2007 and the 2001 turmoil are well captured by the high-volatility regime, while the period starting from the mid 1990s was in the low-volatility regime, sometimes referred to as the big moderation. The macro parameter estimates, annual rates of time preference ρ about 1.4%, and capital depreciation rate about 2.6% with variance σ about 2.3%, are roughly in line with our baseline MEF results in column 1.

The latent variable extension with stochastic volatility of the short rate, case (iii), shows strong mean reversion κ_2 in (the logarithm of) the short rate volatility. The speed of mean reversion, κ_2, is about 280%, such that η_t rapidly reverts back to a value of about 0.006 (which corresponds to γ_2 about -10.4), similar to our baseline estimate (column 1). Based on a regression for the RV proxy series, the volatility (of the logarithm) of the latent short rate volatility, ξ, is identified at about 3.8. Regarding the macro parameter estimates, the estimates for ρ and σ are similar to those in the regime-switching model (column 2), whereas the depreciation rate is somewhat higher, about 7.5%, but still in line with the usual estimates.

The mixed-frequency data analysis, case (iv), indicates that the MF-MEF long-term value for the interest rate γ is even smaller, about 3%, but quite persistent. The point estimate for the speed of mean-reversion parameter κ is 2.3%, which suggests a near unit root behavior. The innovation variance η is close to zero. Our MF-MEF approach yields a more precise estimate for ρ of about 1.3% compared to the MEF estimate of 2.1% and 0.3% for quarterly respectively monthly data. The point estimate for the depreciation rate δ using quarterly GDP and monthly consumption is smaller than the MEF values, about 1% only.

6 Conclusion

The literature has been relatively quiet on the links between macroeconomics and finance, though anecdotal evidence – such as the recent financial crisis – clearly shows that financial markets and the real economy are closely linked. In this paper we provide an econometric framework in which macroeconomics, finance and econometrics are coherently linked. The framework is developed in a continuous-time setting that conveniently allows for thinking about variables observed at different frequencies.

This paper describes various methods, including the GMM and the more efficient MEF approach, in order to estimate the structural parameters of continuous-time DSGE models using mixed-frequency macro and financial market data. We illustrate our approach by solving and estimating a stochastic AK model with mean-reverting interest rates. Our results for both simulated and empirical data are very promising and show that financial and macro data can indeed be used jointly to facilitate the estimation of structural parameters in continuous-time versions of the general equilibrium models. Overall, on the methodological side, our work suggests that MEF is preferred over GMM and regression-based approaches, particularly when the econometrician is restricted to first moments. It allows identifying all structural parameters already from first moments, and estimates are more precise, numerically stable, and economically meaningful. We provide extensions of MEF akin to filtering in order to estimate models with latent variables based on simulations, SMEF, and model-consistent prediction for mixed-frequency data, MF-MEF. Development of further general equilibrium models in the Cox, Ingersoll, and Ross (1985a) framework to more elaborate specifications and formal testing of these is part of our research agenda.

7 Appendix

7.1 The Bellman equation and the Euler equation

As a necessary condition for optimality in our baseline model (cf. Section 2.1), Bellman's principle gives at time s

$$\rho V(K_s, A_s) = \max_{C_s} \left\{ u(C_s, A_s) + \frac{1}{dt} E_s dV(K_s, A_s) \right\}.$$

Itô's formula yields

$$\begin{aligned} dV &= V_K dK_s + V_A dA_s + \tfrac{1}{2}\left(V_{AA}\eta(A_s)^2 + V_{KK}\sigma^2 K_s^2\right) dt \\ &= ((r_s - \delta)K_s + w_s - C_s)V_K dt + V_K \sigma K_s dZ_s + V_A \mu(A_t)dt + V_A \eta(A_s)dB_s \\ &\quad + \tfrac{1}{2}\left(V_{AA}\eta(A_s)^2 + V_{KK}\sigma^2 K_s^2\right) dt. \end{aligned}$$

Using the properties of stochastic integrals, we may write

$$\begin{aligned}\rho V(K_s, A_s) &= \max_{C_s} \{u(C_s, A_s) + ((r_s - \delta)K_s + w_s - C_s)V_K \\ &\qquad + \tfrac{1}{2}\left(V_{AA}\eta(A_s)^2 + V_{KK}\sigma^2 K_s^2\right) + V_A\mu(A_s)\}\end{aligned}$$

for any $s \in [0, \infty)$. Because it is a necessary condition for optimality, we obtain the first-order condition (8), which makes optimal consumption a function of the state variables.

For the *evolution of the costate* we use the maximized Bellman equation

$$\begin{aligned}\rho V(K_t, A_t) &= u(C(K_t, A_t), A_t) + ((r_t - \delta)K_t + w_t - C(K_t, A_t))V_K \\ &\quad + \tfrac{1}{2}\left(V_{AA}\eta(A_t)^2 + V_{KK}\sigma^2 K_t^2\right) + V_A\mu(A_t), \end{aligned} \tag{49}$$

where $r_t = r(K_t, A_t) = Y_K$ and $w_t = w(K_t, A_t) = Y_L$, to compute the costate,

$$\begin{aligned}\rho V_K &= ((r_t - \delta)K_t + w_t - C_t)V_{KK} + (r_t - \delta)V_K \\ &\quad + \tfrac{1}{2}\left(V_{AAK}\eta(A_t)^2 + V_{KKK}\sigma^2 K_t^2\right) + V_{KK}\sigma^2 K_t + V_{AK}\mu(A_t).\end{aligned}$$

Collecting terms we obtain

$$\begin{aligned}(\rho - (r_t - \delta))V_K &= ((r_t - \delta)K_t + w_t - C_t)V_{KK} \\ &\quad + \tfrac{1}{2}\left(V_{AAK}\eta(A_t)^2 + V_{KKK}\sigma^2 K_t^2\right) + V_{KK}\sigma^2 K_t + V_{AK}\mu(A_t).\end{aligned} \tag{50}$$

Using Itô's formula, the costate obeys

$$\begin{aligned}dV_K &= V_{AK}\mu(A_t)dt + V_{AK}\eta(A_t)dB_t \\ &\quad + \tfrac{1}{2}\left(V_{KAA}\eta(A_t)^2 + V_{KKK}\sigma^2 K_t^2\right)dt \\ &\quad + ((r_t - \delta)K_t + w_t - C_t)V_{KK}dt + V_{KK}\sigma K_t dZ_t,\end{aligned}$$

where inserting (50) into the last expression yields

$$dV_K = (\rho - (r_t - \delta))V_K dt - V_{KK}\sigma^2 K_t dt + V_{AK}\eta(A_t)dB_t + V_{KK}\sigma K_t dZ_t,$$

which describes the evolution of the costate variable. As a final step, we insert the first-order condition (8) to obtain the Euler equation (9).

As shown in Posch (2009), the model has a closed-form solution, and the value function is $V(K_t, A_t) = \ln K_t/\rho + f(A_t)$, where $f(A_t)$ solves a simple ODE, which in turn depends on the functional forms of $\eta(A_t)$ and $\mu(A_t)$. The idea of this proof is as follows. We use a guess for the value function and obtain conditions under which both the maximized Bellman equation (49) and the first-order condition (8) are fulfilled. Our guess is

$$V(K_t, A_t) = \mathbb{C}_1 \ln K_t + f(A_t). \tag{51}$$

From (8), optimal consumption is a constant fraction of wealth, $C_t = \mathbb{C}_1^{-1} K_t$. Now use the maximized Bellman equation (49) and insert the candidate solution,

$$\rho \mathbb{C}_1 \ln K_t + g(A_t) = \ln K_t - \ln \mathbb{C}_1 + ((A_t - \delta) K_t - \mathbb{C}_1^{-1} K_t) \mathbb{C}_1 / K_t,$$

in which $g(A_t) \equiv \rho f(A_t) - \frac{1}{2}\left(f_{AA}\eta(A_t)^2 - \sigma^2\right) - f_A \mu(A_t)$. Thus, we obtain the condition $\mathbb{C}_1 = 1/\rho$ and collect the remaining terms in $g(A_t) = \ln \rho + A_t - \delta - \rho$. In the Vasicek case, $\eta(A_t) = \eta$ and $\mu(A_t) = \kappa(\gamma - A_t)$, we get $f(A_t) = \mathbb{C}_2 A_t + \mathbb{C}_3$, in which $\mathbb{C}_2 = \mathbb{C}_1/(\rho + \kappa)$ and $\mathbb{C}_3 = (\kappa\gamma\mathbb{C}_2 - \ln \mathbb{C}_1 - 1 - (\delta + \frac{1}{2}\sigma^2)\mathbb{C}_1)/\rho$.

7.2 Proof of Theorem 3.2

(a) and (b) are among the results of Hansen (1982). The remaining results follow along the lines of Godambe and Heyde (1987), Heyde (1997), and Christensen and Sørensen (2008). In particular, any estimating function given on the form $M_T = \sum_{t=1}^{T} w_t m_t$ as in (26), with w_t in the information set at $t-1$, is a zero-mean martingale. This follows because

$$\begin{aligned} E_{T-1}(M_T) &= \sum_{t=1}^{T-1} w_t m_t + w_T E_{T-1}(m_T) \\ &= M_{T-1}, \end{aligned} \tag{52}$$

using the conditional moment restrictions $E_{T-1}(m_T) = 0$. Specifically, by iterated expectations,

$$\begin{aligned} E(M_T) &= E(E_{T-1}(M_T)) \\ &= E(M_{T-1}), \end{aligned} \tag{53}$$

so that $E(M_T) = E(M_1) = E(w_1 m_1) = w_1 E(m_1) = 0$. Starting now with Theorem 3.2 (d), it follows from Theorem 2.1 in Heyde (1997) that in the class of estimating functions given on this form, M_T^* given by the choice $w_t = w_t^*$ is optimal in the sense of smallest possible asymptotic variance if and only if the matrix

$$\left(E\left(\frac{\partial M_T}{\partial \phi^\top}\right)\right)^{-1} E\left(M_T (M_T^*)^\top\right) \tag{54}$$

is the same for all estimating functions M_T in the class. We have

$$\begin{aligned} E\left(\frac{\partial M_T}{\partial \phi^\top}\right) &= \sum_{t=1}^{T} E\left(w_t \frac{\partial m_t}{\partial \phi^\top}\right) \\ &= \sum_{t=1}^{T} E\left(w_t E_{t-1}\left(\frac{\partial m_t}{\partial \phi^\top}\right)\right) \\ &= \sum_{t=1}^{T} E(w_t \psi_t), \end{aligned} \tag{55}$$

using iterated expectations and the definition (31). Similarly,

$$\begin{aligned}
E\left(M_T\left(M_T^*\right)^\top\right) &= \sum_{s=1}^{T}\sum_{t=1}^{T} E\left(w_s m_s m_t^\top \left(w_t^*\right)^\top\right) \\
&= \sum_{t=1}^{T} E\left(w_t E_{t-1}\left(m_t m_t^\top\right)\left(w_t^*\right)^\top\right) + \sum_{s=1}^{T-1}\sum_{t=s+1}^{T} E\left(w_s m_s E_{t-1}\left(m_t^\top\right)\left(w_t^*\right)^\top\right) \\
&\quad + \sum_{s=2}^{T}\sum_{t=1}^{s-1} E\left(w_s E_{s-1}\left(m_s\right) m_t^\top \left(w_t^*\right)^\top\right) \\
&= \sum_{t=1}^{T} E\left(w_t \Psi_t \left(w_t^*\right)^\top\right), \qquad (56)
\end{aligned}$$

using iterated expectations, the conditional moment conditions, and the definition (30). It follows that when $w_t^* = \psi_t^\top \left(\Psi_t\right)^{-1}$ as in Definition 3.1 (d), expressions (55) and (56) coincide, so that the matrix (54) that is required common across M_T equals the identity matrix, and optimality of the estimating function M_T^* follows.

Theorem 3.2 (c) follows in the usual way from the mean value theorem,

$$0 = M_T^*(\hat{\phi}) = M_T^*(\phi) + S_T(\hat{\phi} - \phi). \qquad (57)$$

Here, the (i,j)'th entry of the $\dim\phi \times \dim\phi$ matrix S_T is

$$\left(S_T\right)_{ij} = \frac{\partial M_T^*(\phi^{(j)})_i}{\partial \phi_j}, \qquad (58)$$

where $\phi^{(j)}$ is a value of the parameter vector on the straight line connecting $\hat{\phi}$ and ϕ. By (55) and ergodicity,

$$\begin{aligned}
\frac{1}{T}\frac{\partial M_T^*(\phi)}{\partial \phi^\top} &\xrightarrow{a.s.} E(w_t^* \psi_t) \\
&= E(\psi_t^\top (\Psi_t)^{-1} \psi_t), \qquad (59)
\end{aligned}$$

where the convergence is with probability one as $T \to \infty$. From (58) and (59),

$$\frac{1}{T} S_T \xrightarrow{a.s.} E(\psi_t^\top (\Psi_t)^{-1} \psi_t), \qquad (60)$$

since the regularity conditions ensure that the convergence is uniform in a $\sqrt{T}$-shrinking neighborhood of ϕ. We have

$$\begin{aligned}
Var\left(\frac{M_T^*(\phi)}{\sqrt{T}}\right) &= \frac{1}{T} E\left(M_T^*\left(M_T^*\right)^\top\right) \\
&= E\left(w_t^* \Psi_t \left(w_t^*\right)^\top\right) \\
&= E(\psi_t^\top (\Psi_t)^{-1} \psi_t), \qquad (61)
\end{aligned}$$

where the first equality is based on the zero mean condition, the second on (56), and the third on the form of the optimal instruments, cf. Definition 3.1 (d). Since $M_T^*(\phi)$ is a zero-mean martingale, and because (61) is regular, so that $Var\left(M_T^*(\phi)\right) \stackrel{a.s.}{\longrightarrow} \infty$, we have by (60) and the strong law for martingales, see Hall and Heyde (1980), that

$$S_T^{-1} M_T^*(\phi) \stackrel{a.s.}{\longrightarrow} 0, \tag{62}$$

and strong consistency of $\hat{\phi}$ follows from (57). By the martingale central limit theorem, see Hall and Heyde (1980),

$$\frac{1}{\sqrt{T}} M_T^*(\phi) \stackrel{d}{\longrightarrow} \mathcal{N}(0, E(\psi_t^\top (\Psi_t)^{-1} \psi_t)), \tag{63}$$

where the form of the asymptotic variance follows from (61). Combining (57), (60), and (63), we have

$$\begin{aligned} \sqrt{T}\left(\hat{\phi}-\phi\right) &= -\left(\frac{S_T}{T}\right)^{-1} \frac{M_T^*(\phi)}{\sqrt{T}} \\ &\stackrel{d}{\longrightarrow} \mathcal{N}(0, \left(E(\psi_t^\top (\Psi_t)^{-1} \psi_t)\right)^{-1} E(\psi_t^\top (\Psi_t)^{-1} \psi_t) \left(E(\psi_t^\top (\Psi_t)^{-1} \psi_t)\right)^{-1}) \\ &\quad = \mathcal{N}(0, \left(E(\psi_t^\top (\Psi_t)^{-1} \psi_t)\right)^{-1}). \end{aligned} \tag{64}$$

Consistent estimation of the asymptotic variance by (37) follows from stationarity and ergodicity.

Theorem 3.2 (g) follows because an estimating function of the form $\sum_{t=1}^T g_t h_t = \sum_{t=1}^T g_t (z_t \otimes m_t)$ may be written as $\sum_{t=1}^T g_t (z_t \otimes I_{\dim m}) m_t = \sum_{t=1}^T w_t m_t$, where $w_t = g_t (z_t \otimes I_{\dim m})$. This implies that using the expanded set of moment conditions based on h_t is a special case of using the original moment conditions based on m_t. In both cases, the instrument matrices (or weights, i.e., g_t respectively w_t) need only have $\dim \phi$ rows, as argued in the main text before Theorem 3.2: Using more rows is GMM, but an estimator asymptotically equivalent to optimal GMM is obtained using constant $g_t = G$ of dimension $\dim \phi \times \dim h$. Since using h_t is a special case of using m_t, the optimal MEF estimator based on the latter is obviously at least as efficient as that based on the former. That the two actually coincide requires showing that given moment conditions based on h_t, it is possible to recover those based on m_t. This follows because any estimating function based on m_t may be recast in terms of h_t by writing

$$\begin{aligned} \sum_{t=1}^T w_t m_t &= \sum_{t=1}^T w_t (z_t^- \otimes I_{\dim m})(z_t \otimes I_{\dim m}) m_t \\ &= \sum_{t=1}^T g_t h_t, \end{aligned} \tag{65}$$

with $g_t = w_t(z_t^- \otimes I_{\dim m})$, where $z_t^- = z_t^\top / z_t^\top z_t$ is a left inverse of z_t, viz., $z_t^- z_t = 1$.

Theorem 3.2 (f) also follows from these calculations, since for GMM to coincide with optimal MEF it is necessary that the h_t in GMM equal $\psi_t^\top (\Psi_t)^{-1} m_t$, which can only happen in the two cases stated.

For Theorem 3.2 (e), that optimal MEF is at least as efficient as GMM follows from combining (d) and (g), including the proof of (g). There is no gain to expanding from moments m_t to h_t, by the proof of (g). Given moment conditions based on m_t, optimal MEF is at least as efficient as any other MEF estimator, by (d), and therefore at least as efficient as GMM based on m_t, as this is a special case of MEF (namely, with constant $g_t \equiv G$).

Finally, to prove the strict inequality, that optimal MEF is strictly more efficient than GMM unless the two estimators coincide, consider again the optimal MEF estimator given as the solution with respect to ϕ of the system $\sum_{t=1}^T w_t^* m_t = 0$, with $w_t^* = \psi_t^\top (\Psi_t)^{-1}$. Now recall that for zero mean variables x and y with finite variances, the mean square error predictor of y given x is $Cov(y,x)Var(x)^{-1}x$, with prediction error variance-covariance matrix given by $Var(y) - Cov(y,x)Var(x)^{-1}Cov(x,y)$. We apply this to the case $y = w_t^* m_t$ and $x = m_t$. In particular,

$$
\begin{aligned}
Var(w_t^* m_t) &= E\left(w_t^* m_t m_t^\top \left(w_t^*\right)^\top\right) \\
&= E\left(w_t^* E_{t-1}\left(m_t m_t^\top\right)\left(w_t^*\right)^\top\right) \\
&= E\left(w_t^* \Psi_t \left(w_t^*\right)^\top\right) \\
&= E(\psi_t^\top (\Psi_t)^{-1} \psi_t),
\end{aligned}
\tag{66}
$$

by iterated expectations. Similarly,

$$
\begin{aligned}
Cov(w_t^* m_t, m_t) &= E\left(w_t^* m_t m_t^\top\right) \\
&= E\left(w_t^* E_{t-1}\left(m_t m_t^\top\right)\right) \\
&= E\left(\psi_t^\top (\Psi_t)^{-1} \Psi_t\right) \\
&= E\left(\psi_t^\top\right) \\
&= E\left(\left(E_{t-1}\left(\frac{\partial m_t}{\partial \phi^\top}\right)\right)^\top\right) \\
&= E\left(\frac{\partial m_t^\top}{\partial \phi}\right),
\end{aligned}
\tag{67}
$$

by repeated use of iterated expctations. Thus, the mean square error predictor of $w_t^* m_t$ given m_t is $p_t = E(\partial m_t^\top / \partial \phi) Var(m_t)^{-1} m_t$, with prediction error variance-covariance matrix given by

$$\begin{aligned} Var(w_t^* m_t - p_t) &= E(\psi_t^\top (\Psi_t)^{-1} \psi_t) - E\left(\frac{\partial m_t^\top}{\partial \phi}\right) Var(m_t)^{-1} E\left(\frac{\partial m_t}{\partial \phi^\top}\right) \\ &= (V_{MEF})^{-1} - (V_{GMM})^{-1}, \end{aligned} \quad (68)$$

using Theorem 3.2 (a) and (c), the former with $h_t = m_t$ (as already argued, we must compare to GMM based on $h_t = m_t$). Here, since the left-hand side is a variance-covariance matrix, it is positive semi-definite, and since the right-hand side is the difference between the MEF and GMM inverse variances (or precisions), we have again the weak inequality $V_{MEF} \leq V_{GMM}$. Furthermore, $V_{MEF} = V_{GMM}$ requires that the right-hand side of (68) vanishes identically. Because the left-hand side is a variance-covariance matrix, it vanishes only if the distribution of the random vector, in this case $w_t^* m_t - p_t$, is degenerate, i.e., $w_t^* m_t = p_t = E(\partial m_t^\top / \partial \phi) Var(m_t)^{-1} m_t$. This requires that the time-varying MEF matrices w_t^* simplify to the constant $w_t^* = E(\partial m_t^\top / \partial \phi) Var(m_t)^{-1}$, i.e., the matrix G shown in the main text before Theorem 3.2 to correspond to optimal GMM (with $h_t = m_t$). Thus, optimal MEF and GMM coincide in this case. In all other cases, the left-hand side in (68) above is strictly positive definite, hence so is the right-hand side, i.e., $V_{MEF} < V_{GMM}$.

References

Aït-Sahalia, Y. (1996): "Nonparametric Pricing of Interest rate Derivative Securities," *Econometrica*, 64(3), 527–560.

——— (2008): "Closed-form Likelihood Expansions for Multivariate Diffusions," *Annals of Statistics*, 36(2), 906–937.

Amemiya, T., and R. Y. Wu (1972): "The Effect of Aggregation on Prediction in the Autoregressive Model," *J. Amer. Statistical Assoc.*, 67(339), 628–632.

An, S., and F. Schorfheide (2007): "Bayesian Analysis of DSGE models," *Econometric Rev.*, 26(2-4), 113–172.

Andersen, T. G., and T. Bollerslev (1998): "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *Int. Econ. Rev.*, 39(4), 885–905.

Andersen, T. G., and J. Lund (1997): "Estimating continuous-time stochastic volatility models of the short-term interest rate," *J. Econometrics*, 77, 343–377.

Andreasen, M. M., and B. J. Christensen (2015): "The SR Approach: A New Estimation Method for Non-Linear and Non-Gaussian Dynamic Term Structure Models," *J. Econometrics*, 184, 420–451.

Andreou, E., E. Ghysels, and A. Kourtellos (2010): "Regression models with mixed sampling frequencies," *J. Econometrics*, 158, 246–261.

——— (2013): "Should Macroeconomic Forecasters Use Daily Financial Data and How?," *J. Bus. Econ. Statist.*, 31, 1–12.

Ang, A., and M. Piazzesi (2003): "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *J. Monet. Econ.*, 50, 745–787.

Ang, A., M. Piazzesi, and M. Wei (2006): "What does the yield curve tell us about GDP growth?," *J. Econometrics*, 131, 359–403.

Barndorff-Nielsen, O. E., and N. Shephard (2002): "Estimating quadratic variation using realised volatility," *J. Appl. Econometrics*, 17, 457–477.

Bates, C. E., and H. White (1993): "Determination of Estimators With Minimum Asymptotic Covariance Matrices," *Econometric Theory*, 9, 633–648.

Bergstrom, A. R. (1966): "Nonrecursive Models as Discrete Approximations to Systems of Stochastic Differential Equations," *Econometrica*, 34(1), 173–182.

Bernanke, B., and M. Gertler (1989): "Agency Costs, Net Worth, and Business Fluctuations," *Amer. Econ. Rev.*, 79(1), 14–31.

Bernanke, B., M. Gertler, and S. Gilchrist (1999): *The Financial Accelerator in a Quantitative Business Cycle Framework.* in J.B. Taylor and M. Woodford, Eds., Handbook of Macroeconomics, Volume 1, Chapter 21, 1341-1393.

Bollerslev, T., and H. Zhou (2002): "Estimating stochastic volatility diffusion using conditional moments of integrated volatility," *J. Econometrics*, 109, 33–65.

Brunnermeier, M. K., and Y. Sannikov (2014): "A Macroeconomic Model with a Financial Sector," *Amer. Econ. Rev.*, 104(2), 379–421.

Canova, F. (2007): *Methods for Applied Macroeconomic Research.* Princeton University Press.

Canova, F., and L. Sala (2009): "Back to square one: Identification issues in DSGE models," *J. Monet. Econ.*, 56, 431–449.

CHAMBERLAIN, G. (1987): "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *J. Econometrics*, 34, 305–334.

CHANG, F.-R. (1988): "The Inverse Optimal Problem: A Dynamic Programming Approach," *Econometrica*, 56(1), 147–172.

CHAPMAN, D. A., J. B. LONG, AND N. D. PEARSON (1999): "Using Proxies for the Short rate: When are Three Months Like an Instant?," *Rev. Finan. Stud.*, 12(4), 763–806.

CHRISTENSEN, B. J., AND N. M. KIEFER (2009): *Economic Modeling and Inference*. Princeton University Press.

CHRISTENSEN, B. J., AND M. SØRENSEN (2008): "Optimal Inference in Dynamic Models with Conditional Moment Restrictions," *CREATES Research Paper*, 51.

CHRISTIANO, L. J., AND M. EICHENBAUM (1987): "Temporal Aggregation and Structural Inference in Macroeconomics," *Carnegie-Rochester Conference Series on Public Policy*, 26, 63–130.

CHRISTIANO, L. J., R. MOTTO, AND M. ROSTAGNO (2014): "Risk Shocks," *Amer. Econ. Rev.*, 104(1), 27–65.

COX, J. C., J. E. INGERSOLL, AND S. A. ROSS (1985a): "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53(2), 363–384.

——— (1985b): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53(2), 385–407.

DE JONG, F. (2000): "Time Series and Cross-Section Information in Affine Term-Structure Models," *J. Bus. Econ. Statist.*, 18(3), 300–314.

DEWACHTER, H., AND M. LYRIO (2006): "Macro Factors and the Term Structure of Interest Rates," *J. Money, Credit, Banking*, 38(1), 119–140.

DIEBOLD, F. X., G. R. RUDEBUSCH, AND S. B. ARUOBA (2006): "The macroeconomy and the yield curve: a dynamic latent factor approach," *J. Econometrics*, 131, 309–338.

EATON, J. (1981): "Fiscal Policy, Inflation and the Accumulation of Risky Capital," *Rev. Econ. Stud.*, 48, 435–445.

FERNÁNDEZ-VILLAVERDE, J., P. GUERRÓN-QUINTANA, AND J. RUBIO-RAMÍREZ (2015): "Estimating dynamic equilibrium models with stochastic volatility," *J. Econometrics*, (185), 216–229.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Rev. Econ. Stud.*, 74, 1059–1087.

Foerster, A., J. Rubio-Ramírez, D. F. Waggoner, and T. Zha (2015): "Perturbation Methods for Markov-Switching DSGE Models," *Working Paper*.

Foroni, C., and M. Marcellino (2014): "Mixed-Frequency Structural Models: Identification, Estimation, and Policy Analysis," *J. Appl. Econometrics*.

French, K., G. Schwert, and R. Stambaugh (1987): "Expected stock returns and volatility," *J. Fin. Econ.*, 19, 3–30.

Gertler, M., and P. Karadi (2011): "A model of unconventional monetary policy," *J. Monet. Econ.*, 58, 17–34.

Geweke, J. (1978): "Temporal Aggregation in the Multiple Regression Model," *Econometrica*, 46(3), 643–661.

Ghysels, E., A. Sinko, and R. Valkanov (2007): "MIDAS Regressions: Further Results and New Directions," *Econometric Rev.*, 26, 53–90.

Giannone, D., F. Monti, and L. Reichlin (2014): "Exploiting the monthly data-flow in structural forecasting," *mimeo*.

Godambe, V. P., and C. C. Heyde (1987): "Quasi Likelihood and Optimal Estimation," *International Statistical Review*, 55, 231–244.

Hall, P., and C. C. Heyde (1980): *Martingale Limit Theory and Its Applications*. Academic Press, New York.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

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

——— (1985): "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *J. Econometrics*, 30, 203–238.

Hansen, L. P., J. C. Heaton, and M. Ogaki (1988): "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *J. Amer. Statistical Assoc.*, 83, 863–871.

Hansen, L. P., and T. J. Sargent (1991): *Rational Expectations Econometrics*. Westview Press, Boulder.

Hansen, L. P., and J. A. Scheinkman (1995): “Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes,” *Econometrica*, 63(4), 767–804.

Harvey, A., and J. H. Stock (1989): “Estimating Integrated Higher-Order Continuous Time Autoregressions with an Application to Money-Income Causality,” *J. Econometrics*, 42, 319–336.

Heyde, C. C. (1997): *Quasi-Likelihood and Its Application*. Springer, New York.

Hördahl, P., O. Tristani, and D. Vestin (2006): “A joint econometric model of macroeconomic and term-structure dynamics,” *J. Econometrics*, 131, 405–444.

Jermann, U. J. (1998): “Asset Pricing in Production Economies,” *J. Monet. Econ.*, 41, 257–275.

Jermann, U. J., and V. Quadrini (2012): “Macroeconomic Effects of Financial Shocks,” *Amer. Econ. Rev.*, 102(1), 238–271.

Kim, T. B. (2010): “Temporal Aggregation Bias and Mixed Frequency Estimation of a New Keynesian model,” *mimeo.*

Kiyotaki, N., and J. Moore (1997): “Credit Cycles,” *J. Polit. Economy*, 105(2), 211–248.

Kuersteiner, G. M. (2001): “Optimal Instrumental Variables Estimation for ARMA Models,” *J. Econometrics*, 104, 359–405.

Liu, Z., D. F. Waggoner, and T. Zha (2011): “Sources of macroeconomic fluctuations: A regime-switching DSGE approach,” *Quantitative Economics*, 2, 251–301.

Liu, Z., P. Wang, and T. Zha (2013): “Land-Price Dynamics and Macroeconomic Fluctuations,” *Econometrica*, 81(3), 1147–1184.

Marcellino, M. (1999): “Some Consequences of Temporal Aggregation in Empirical Analysis,” *J. Bus. Econ. Statist.*, 17(1), 129–136.

McCrorie, J. R. (2009): “Estimating Continuous-Time Models on the Basis of Discrete Data via an Exact Discrete Analog,” *Econometric Theory*, 25, 1120–1137.

MERTON, R. (1980): "On estimating the expected return on the market," *J. Fin. Econ.*, 8, 323–361.

NEWEY, W. (1990): "Semiparametric Efficiency Bounds," *J. Appl. Econometrics*, 5, 99–135.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

PHILLIPS, P. C. B. (1972): "The Structural Estimation of a Stochastic Differential Equation System," *Econometrica*, 40(6), 1021–1041.

——— (1991): "Error-Correction and Long-Run Equilibrium in Continuous-Time," *Econometrica*, 59(4), 967–980.

POSCH, O. (2009): "Structural Estimation of Jump-Diffusion Processes in Macroeconomics," *J. Econometrics*, 153(2), 196–210.

——— (2011): "Risk premia in general equilibrium," *J. Econ. Dynam. Control*, 35(9), 1557–1576.

PROTTER, P. E. (2004): *Stochastic Integration and Differential Equations*. Springer, Berlin.

ROBINSON, P. M. (1991): "Best Nonlinear Three-Stage Least Squares Estimation of Certain Econometric Models," *Econometrica*, 59, 755–786.

RUDEBUSCH, G. D., AND E. T. SWANSON (2008): "Examining the Bond Premium Puzzle with a DSGE Model," *J. Monet. Econ.*, 55, S111–S126.

RUDEBUSCH, G. D., AND T. WU (2008): "A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy," *Econ. J.*, 118, 906–926.

RUGE-MURCIA, F. J. (2007): "Methods to estimate dynamic stochastic general equilibrium models," *J. Econ. Dynam. Control*, 31, 2599–2636.

SCHORFHEIDE, F., AND D. SONG (2014): "Real-Time Forecasting with a Mixed-Frequency VAR," *J. Bus. Econ. Statist.*, forthcoming.

SENNEWALD, K. (2007): "Controlled Stochastic Differential Equations under Poisson Uncertainty and with Unbounded Utility," *J. Econ. Dynam. Control*, 31(4), 1106–1131.

SIMS, C. A. (1971): "Discrete Approximations to Continuous Time Distributed Lags in Econometrics," *Econometrica*, 39(3), 545–563.

Sims, C. A., D. F. Waggoner, and T. Zha (2008): "Methods for inference in large multiple-equation Markov-switching models," *J. Econometrics*, 146, 255–274.

Tallarini, T. (2000): "Risk-Sensitive Real Business Cycles," *J. Monet. Econ.*, 45, 507–532.

Tang, C. Y., and S. X. Chen (2009): "Parameter estimation and bias correction for diffusion processes," *J. Econometrics*, 149, 65–81.

van Binsbergen, J. H., J. Fernández-Villaverde, R. S. J. Koijen, and J. F. Rubio-Ramírez (2012): "The Term Structure of Interest Rates in a DSGE Model with Recursive Preferences," *J. Monet. Econ.*, 59, 634–648.

Vasicek, O. (1977): "An Equilibrium Characterization of the Term Structure," *J. Finan. Econ.*, 5, 177–188.

Wang, X., P. C. B. Phillips, and J. Yu (2011): "Bias in estimating multivariate and univariate diffusions," *J. Econometrics*, 161, 228–245.

Yu, J. (2012): "Bias in the estimation of the mean reversion parameter in continuous time models," *J. Econometrics*, 169, 114–122.

Zhang, L., P. Mykland, and Y. Aït-Sahalia (2005): "A tale of two time scales: Determining integrated volatility with noisy high-frequency data," *J. Amer. Statistical Assoc.*, 100, 1394–1411.

Table 1: Simulation Study – AK-Vasicek with Three Conditional Moment Restrictions

The table reports output of a simulation study of the accuracy of the structural model parameters estimated using the OLS, FGLS-SUR-IV, GMM and MEF approaches for the AK-Vasicek model. For 1,000 replications, we generate 25 years of data from the underlying data generating process (DGP) and apply our estimation strategy. We show the median estimate, and provide the interquartile range below it.

Parameter Estimates from Simulation Study – Monthly & Quarterly Data									
		Monthly Data				Quarterly Data			
	DGP	OLS	FGLS-SUR-IV	GMM	MEF	OLS	FGLS-SUR-IV	GMM	MEF
κ	0.2	0.349	0.299	0.345	0.354	0.354	0.225	0.287	0.353
		0.286	0.134	0.345	0.284	0.290	0.119	0.319	0.305
γ	0.1	0.201	0.101	0.100	0.099	0.198	0.100	0.101	0.099
		0.036	0.013	0.014	0.013	0.036	0.014	0.015	0.013
η	0.01	0.083	0.008	0.010	0.010	0.083	0.007	0.010	0.010
		0.036	0.004	0.001	0.001	0.035	0.003	0.002	0.001
ρ	0.03	0.080	0.030	0.030	0.030	0.079	0.030	0.031	0.030
		0.015	0.006	0.007	0.006	0.015	0.006	0.007	0.006
δ	0.05	0.05	0.05	0.05	0.050	0.05	0.05	0.05	0.050
					0.002				0.003
σ	0.02	0.317	0.000	0.027	0.023	0.312	0.000	0.040	0.025
		0.040	<0.001	0.047	0.005	0.044	<0.001	0.064	0.010

Table 2: Estimates – AK-Vasicek with Three Conditional Moment Restrictions

The table reports estimates for the structural model parameters estimated using OLS, FGLS-SUR-IV, GMM, and MEF approaches for the AK-Vasicek model with three conditional moment restrictions. We run the estimation for monthly data (where production is measured by IP) and quarterly data (production measured by GDP). The sample runs from January, 1982, through December, 2012. Asymptotic t-statistics are given below the estimates.

Parameter Estimates from Empirical Data								
	Monthly Data				Quarterly Data			
	OLS	FGLS-SUR-IV	GMM	MEF	OLS	FGLS-SUR-IV	GMM	MEF
κ	0.094	0.073	0.054	0.060	0.109	0.047	0.062	0.068
	0.582	2.46	0.314	0.930	1.94	0.221	0.971	0.515
γ	0.094	0.089	0.064	0.051	0.129	0.119	0.095	0.131
	2.08	$<$0.001	0.549	0.479	$<$0.001	7.97	2.25	0.483
η	0.016	0.013	0.000	0.006	0.026	0.013	0.000	0.014
	0.097	$<$0.001	$<$0.001	0.639	0.081	1.97	$<$0.001	0.221
ρ	0.014	0.014	0.005	0.003	0.021	0.020	0.015	0.021
	0.976	0.430	0.129	0.802	0.974	0.871	0.753	5.95
δ	0.05	0.05	0.05	0.025	0.05	0.05	0.05	0.062
				0.162				0.144
σ	0.02	0.02	0.02	0.020	0.02	0.02	0.02	0.020
				3.3				1.15

Table 3: Estimates – MEF Extensions

The table reports estimates for the structural model parameters for the MEF extensions: The latent short rate and mixed-frequency approaches for the AK-Vasicek model, SMEF (Latent Short Rate) and MF-MEF, respectively, and for the regime-switching (RS) and stochastic volatility (SV) models. We run the estimation for monthly data (where production is measured by IP), quarterly data (production measured by GDP) and the mixed-frequency data (where production is quarterly GDP data, and monthly consumption data). The sample runs from January, 1982, through December, 2012. Asymptotic t-statistics are given below the estimates.

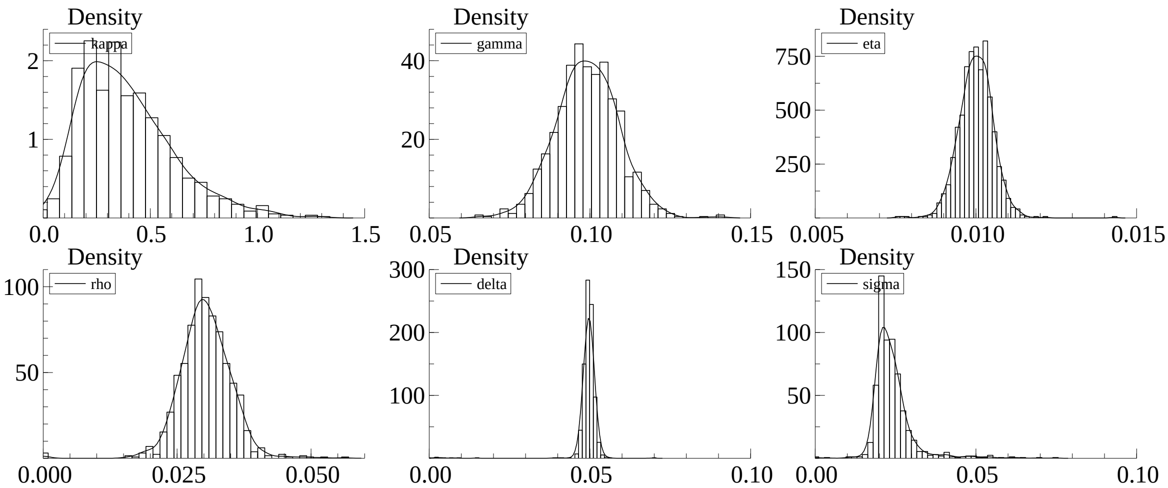
Parameter Estimates from Empirical Data						
	Monthly Data			Quarterly Data		Mixed-Frequency
	MEF	RS	SV	MEF	SMEF	MF-MEF
κ/κ_1	0.060 0.93	0.006 0.152	0.073 0.069	0.068 0.515	0.126 3.2	0.023 0.137
γ/γ_1	0.051 0.479	0.125 0.3	0.087 0.045	0.131 0.483	0.058 55.8	0.030 0.937
η	0.006 0.639			0.014 0.221	0.000 0.3	0.001 0.005
η_l		0.005 11.2^a				
η_h		0.021 10.0^a				
ϕ_{lh}		0.591 2.36^a				
ϕ_{hl}		1.464 2.22^a				
κ_2			2.818 34.8			
γ_2			−10.377 −251.4			
ξ			3.834 0.707^b			
ρ	0.003 0.802	0.014 3.524	0.012 0.012	0.021 5.95	0.009 4.39	0.013 23.2
δ	0.025 0.162	0.026 0.475	0.075 0.165	0.062 0.144	0.024 12.0	0.008 0.042
σ	0.020 3.3	0.023 9.116	0.021 0.604	0.020 1.15	0.022 510	0.025 1.06

[a] Based on Hamilton filter for interest rate only
[b] Based on regression for proxy series

Figure 1: Simulation Study – Monthly and Quarterly Data

The figure reports output of a simulation study of the accuracy of the structural model parameters estimated using the MEF approach for the AK-Vasicek model. For 1,000 replications, we generate 25 years of data from the underlying data generating process (DGP) and apply our estimation strategy. We plot the distribution of the obtained estimates, in Panel A for monthly data and in Panel B for quarterly data.

(A) Monthly Data



(B) Quarterly Data

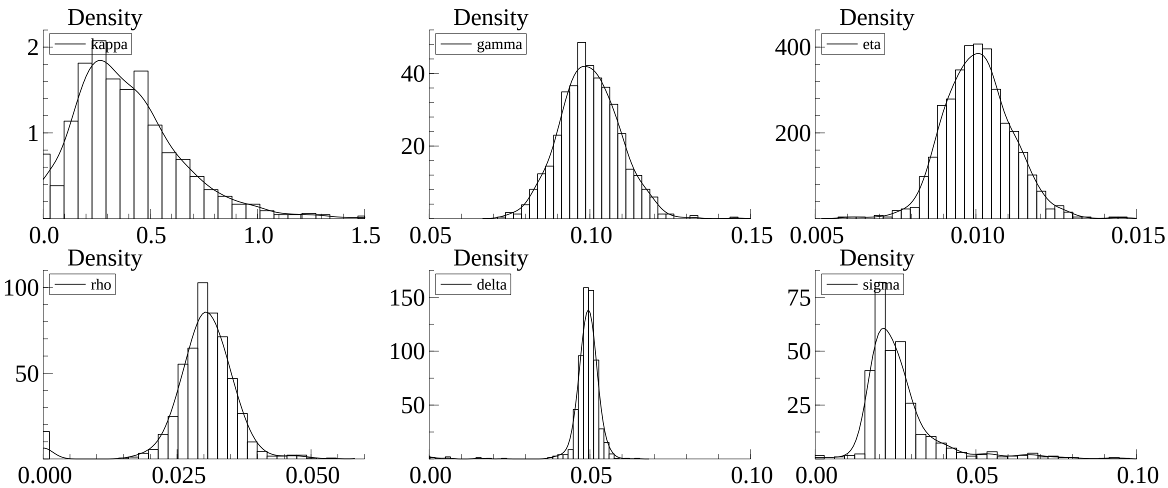
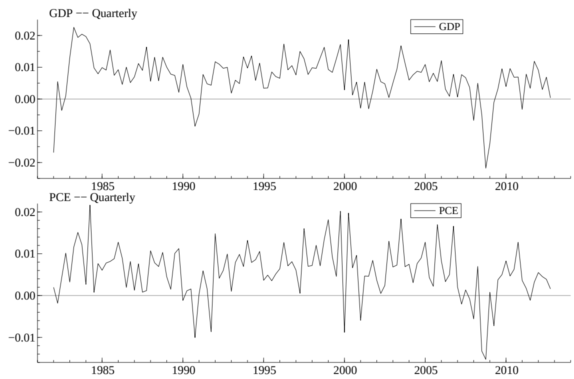


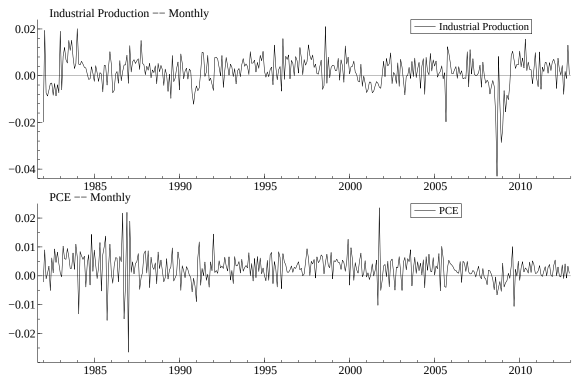
Figure 2: Overview of Quarterly, Monthly and Daily Variables

The figure shows time series plots of the variables in our data set at the quarterly (Panel A), monthly (Panel B), and daily (Panel C) frequency. In Panel A, the top plot shows the growth rate of real Gross Domestic Product (GDP), and the bottom plot that of real Personal Consumption Expenditure (PCE), both at the quarterly frequency. In Panel B, the top plot shows the growth rate of Industrial Production (IP), and the bottom plot that of real PCE, both at the monthly frequency. Panel C shows the nominal 3m interest rate series at the daily frequency. All series are obtained from the Federal Reserve Bank of St. Louis Economic Dataset (FRED). The sample runs from January, 1982, through December, 2012.

(A) Quarterly Macroeconomic Variables



(B) Monthly Macroeconomic Variables



(C) Daily Interest Rate

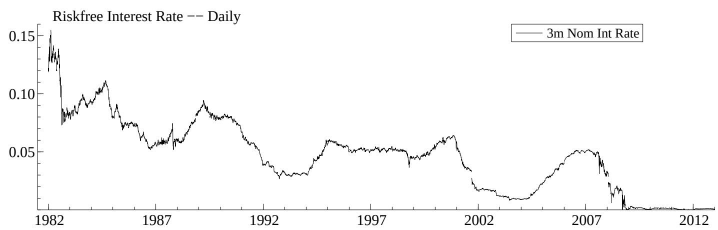
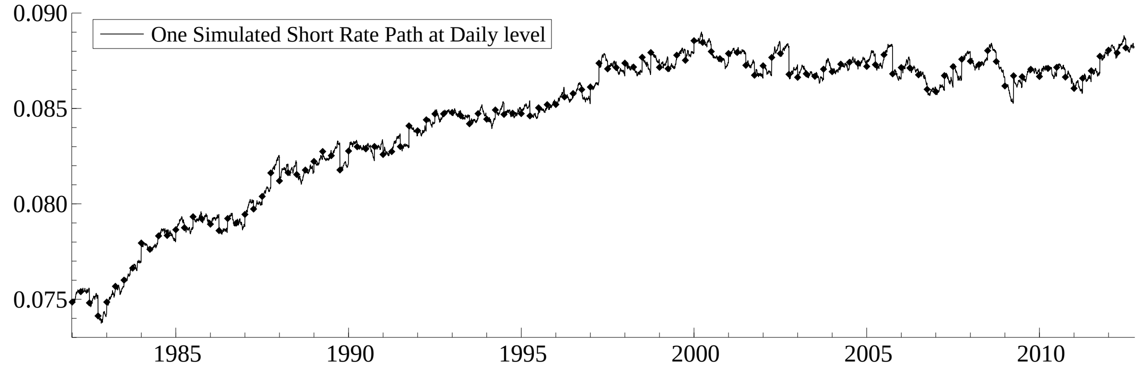


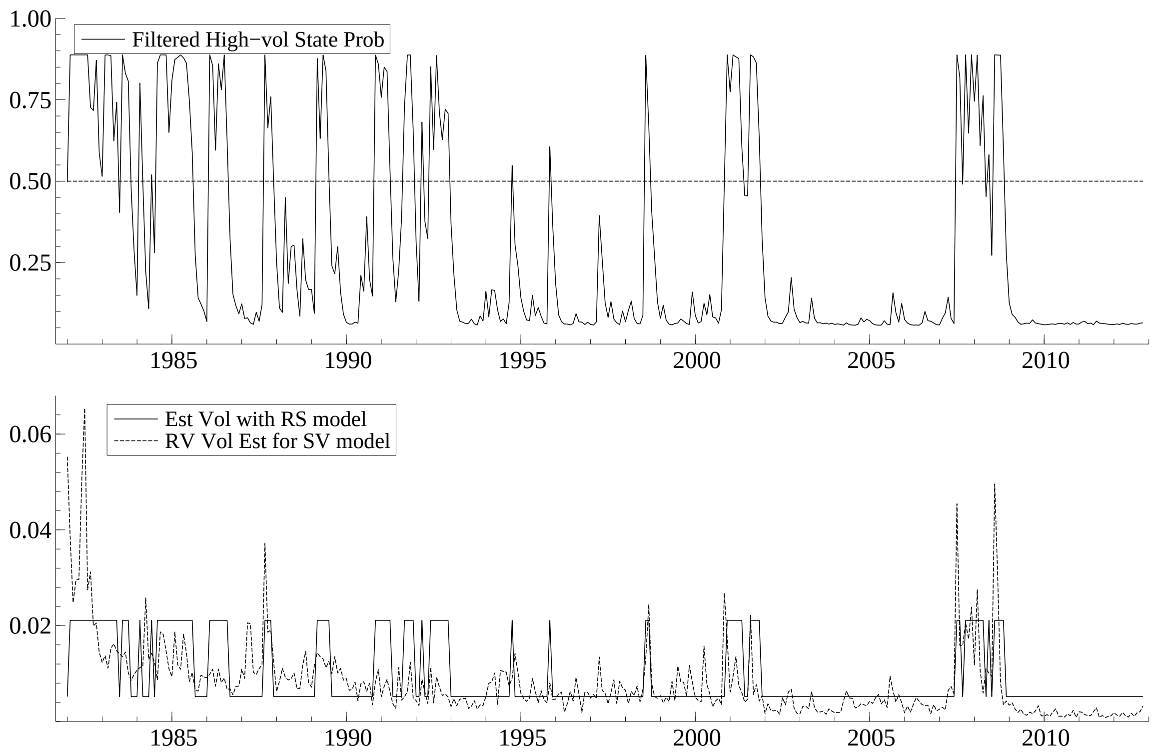
Figure 3: Simulated Latent Short Rate and Estimated Volatility Regimes

In this figure we show the simulated latent short rate and the output from the regime-switching model. Panel (A) provides the simulated latent short rate that is obtained by applying our latent short rate extended MEF approach to the AK-Vasicek model. We report the short rate estimate based on quarterly data. The line represents one path of the simulated short rate series, and the dots the proxy $r_t^* = \rho Y_t / C_t$. Panel (B) reports the output of the Regime Switching (RS) model estimated using the regime-switching extended MEF approach for monthly data. The top plot provides the estimated probability of being in the high-volatility state (solid line; dashed line is 50% for reference). The bottom plot provides the estimated volatility based on the Regime Switching model (solid line), along with the Realized Volatility (RV) estimate based on daily short rate data for comparison (dashed line). In both panels, the sample runs from January, 1982, through December, 2012.

(A) Simulated Short Rate



(B) Regime Switching Output



Estimating Dynamic Equilibrium Models using Mixed Frequency Macro and Financial Data

Web Appendix

Bent Jesper Christensen[(a,b)], Olaf Posch[(c,b)], and Michel van der Wel[(b,d)]
[(a)]Aarhus University, [(b)]CREATES, [(c)] Universität Hamburg,
[(d)]Erasmus University Rotterdam

April 18, 2016

Contents

A Alternative estimation approaches

A.1 The regression-based approaches

In this section we propose regression-based procedures to obtain benchmark parameter estimates. To start with, we employ unrestricted ordinary least squares (OLS) to get reduced-form parameters, although this does not identify the structural parameters of interest. Next, we consider cross-equation correlation, controlling for endogeneity through instrumental variables (IV), and estimation of structural parameters by minimum distance.

A.1.1 Reduced-form model

With $s-t$ fixed at Δ, and using the proxy series $\hat{r}_t$ in (17), the system (15) is linear in a set of reduced-form parameters and may be recast as

$$y_{j,t} = x_{j,t}\beta_j + \varepsilon_{j,t}, \qquad j = C, Y, r, \tag{A.1}$$

where the left-hand side variables are $y_{C,t} = \ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^{t} r_v^f dv$, $y_{Y,t} = \ln(Y_t/Y_{t-\Delta}) - \int_{t-\Delta}^{t} r_v^f dv$, and $y_{r,t} = r_t^f$.[1] The right-hand side variables $x_t = (x_{C,t}, x_{Y,t}, x_{r,t})$, with $x_{C,t} = 1$, $x_{Y,t} = (1, \int_{t-\Delta}^{t} 1/\hat{r}_v dv, \int_{t-\Delta}^{t} 1/\hat{r}_v^2 dv)$, and $x_{\hat{r},t} = (1, \hat{r}_{t-\Delta})$. The reduced-form or linear parameters, β_C, $\beta_Y = (\beta_{Y,1}, \beta_{Y,2}, \beta_{Y,3})^\top$, and $\beta_r = (\beta_{r,1}, \beta_{r,2})^\top$, are given in terms of the structural parameters $\phi = (\kappa, \gamma, \eta, \rho, \delta, \sigma)^\top$ as

$$\beta_C = -\left(\rho - \tfrac{1}{2}\sigma^2\right)\Delta, \tag{A.2a}$$

$$\begin{aligned} \beta_{Y,1} &= -\left(\kappa + \rho - \tfrac{1}{2}\sigma^2\right)\Delta, \\ \beta_{Y,2} &= \kappa\gamma, \\ \beta_{Y,3} &= -\tfrac{1}{2}\eta^2, \end{aligned} \tag{A.2b}$$

$$\begin{aligned} \beta_{r,1} &= (1 - e^{-\kappa\Delta})(\gamma - \delta - \sigma^2), \\ \beta_{r,2} &= e^{-\kappa\Delta}. \end{aligned} \tag{A.2c}$$

Hence, the system (15) can be summarized in the form of simple regression equations, with error terms given by

$$\varepsilon_{C,t} = \sigma(Z_t - Z_{t-\Delta}), \tag{A.3a}$$

$$\varepsilon_{Y,t} = \int_{t-\Delta}^{t} \eta/\hat{r}_v dB_v + \sigma(Z_t - Z_{t-\Delta}), \tag{A.3b}$$

$$\varepsilon_{r,t} = \eta e^{-\kappa\Delta} \int_{t-\Delta}^{t} e^{\kappa(v-(t-\Delta))} dB_v. \tag{A.3c}$$

[1] In cases where δ and σ in (17) are identified by the remaining system of equations, we may interpret fixed values δ_0 and σ_0 in the construction of the auxiliary variable $\hat{r}_t$ as starting values, then estimate the full set of parameters of the model and update the values for $\delta_i = \hat{\delta}_{i-1}$ and $\sigma_i = \hat{\sigma}_{i-1}$ recursively for $i = 1, 2, \ldots$ Alternatively, a nonlinear one-step regression-based approach could be implemented.

Using iterated expectations and the properties of stochastic integrals, if the parameters are at their true values, including δ_0 and σ_0 in (17), then the error terms are clearly serially uncorrelated, i.e., $E(\varepsilon_{j,t}\varepsilon_{j,t-\Delta}) = 0$, $j = C, Y, r$. For a simple reduced-form estimator, linearity in β suggests unrestricted equation-by-equation OLS,

$$\hat{\beta}_j = (x_j^\top x_j)^{-1} x_j^\top y_j, \qquad j = C, Y, r, \tag{A.4}$$

where x_j is the matrix with typical row $x_{j,t}$ and y_j the vector with typical entry $y_{j,t}$. The structural parameter estimates, obtained by minimum distance applied to the reduced-form estimates (A.4) using the link (A.2a)-(A.2c) (or by an asymptotically equivalent restricted nonlinear least squares regression), serve as useful benchmarks for assessing more elaborate structural approaches. We next discuss enhancing the basic OLS-based estimators by correction for contemporaneous cross-equation correlation of errors and endogeneity of right-hand side variables, then present the minimum distance approach yielding the structural parameter estimates.

A.1.2 Cross-equation correlation

The estimators (A.4) allow for different variances of the error terms $\varepsilon_{j,t}$, say, σ_j^2, $j = C, Y, r$, as they are implemented separately by equation. However, they do not exploit all other properties of the errors. The present model structure implies both different right-hand side variables (indeed, of different dimensions) across the equations, and cross-equation correlation of the errors. In particular, from (A.3a)-(A.3c), the term $\sigma(Z_t - Z_{t-\Delta})$ is common to both $\varepsilon_{C,t}$ and $\varepsilon_{Y,t}$, whereas both $\varepsilon_{Y,t}$ and $\varepsilon_{r,t}$ involve stochastic integrals with respect to B_v. Classical seemingly unrelated regressions (SUR) analysis is intended to exploit such cross-equation correlation in the errors to improve efficiency in estimation exactly in cases where the right-hand side variables are not common across equations. This suggests that a standard SUR correction of the reduced-form estimates should be more efficient than the OLS estimates, and, hence, that structural parameter estimates backed out from the SUR estimates (using minimum distance) should dominate those based on OLS.

Let $\hat{\varepsilon}$ be the $T \times 3$ matrix of OLS residuals, with typical row $(\hat{\varepsilon}_{C,t}, \hat{\varepsilon}_{Y,t}, \hat{\varepsilon}_{r,t})$, where T is the number of time periods in the data set. The SUR estimate of the 3×3 contemporaneous system variance-covariance matrix is $\hat{\Sigma} = \hat{\varepsilon}^\top \hat{\varepsilon}/T$ (in particular, the residual variance estimates along the diagonal coincide with the standard OLS assessments), and the FGLS-SUR estimate of $\beta = (\beta_C, \beta_Y^\top, \beta_r^\top)^\top$ is

$$\hat{\beta}_{SUR} = (x^\top \hat{V}^{-1} x)^{-1} x^\top \hat{V}^{-1} y, \tag{A.5}$$

where y is the $3T$-vector stacking the y_j, x is the conformable matrix with the x_j along the block-diagonal, and $\hat{V}^{-1} = \hat{\Sigma}^{-1} \otimes I_T$, with I_T the identity matrix and $\otimes$ the Kronecker product. The variance-covariance matrices of the SUR (and OLS) estimators are given in Appendix A.2.

A.1.3 Endogeneity

The regression approaches (OLS and SUR) do not control for possible endogeneity of right-hand side variables in (15), and hence (A.1), which may be an issue in the DSGE model. In particular, $x_{Y,t}$ includes two integrals involving the evolution of the auxiliary variable in (17) from $t-\Delta$ through t and so is correlated with both $\varepsilon_{r,t}$ and $\varepsilon_{Y,t}$. The standard regression-based tool for handling endogeneity is instrumental variables (IV). Here, we consider first-stage regressions of each of $x_{Y,t,2} = \int_{t-\Delta}^{t} 1/\hat{r}_v dv$ and $x_{Y,t,3} = \int_{t-\Delta}^{t} 1/\hat{r}_v^2 dv$ on their respective lags $x_{Y,t-\Delta,2}$ and $x_{Y,t-\Delta,3}$ and an intercept. Next, in the computation (A.4) of $\widehat{\beta}_Y$, fitted values from the first stage regressions replace $x_{Y,t,2}$ and $x_{Y,t,3}$. Third, fitted residuals are calculated using the new second stage estimate $\widehat{\beta}_Y$ but the original $x_{Y,t,2}$ and $x_{Y,t,3}$ (not their fitted values from the first stage), and these residuals form the basis of the IV assessment of $\hat{\Sigma}$. Finally, an FGLS-SUR-IV step is carried out using this new $\hat{\Sigma}$ in calculating $\hat{\beta}_{SUR}$ in (A.5) and again using the fitted values for $x_{Y,t,2}$ and $x_{Y,t,3}$. This combination of FGLS, SUR, and IV (labeled FGLS-SUR-IV) appears to be novel.

Note that the lagged values of the relevant integrals involving the auxiliary variable $\hat{r}_s$, $t-2\Delta \leq s \leq t-\Delta$, may correlate with $\hat{r}_{t-\Delta}$, and hence with $\varepsilon_{Y,t}$ from (A.3b), although presumably less than without lagging (this is the idea of the instrumentation). Any such correlation between the error terms and the right-hand side variables (even when using fitted values) indicates that part of the endogeneity issue remains. For a full solution and a consistent and asymptotically efficient estimator, we therefore present the MEF approach in the main text, exploiting the martingale structure of the model.

A.1.4 Minimum distance

The structural parameters are $\phi = (\kappa, \gamma, \eta, \rho, \delta, \sigma)^\top$, a total of six. They are identified by exploiting the way in which they enter into the reduced-form parameters $\beta = \beta(\phi)$. From (A.2a)-(A.2c), we may this way identify $\rho - \frac{1}{2}\sigma^2$, κ, γ, η, and $\delta + \sigma^2$, i.e., three structural parameters, and two independent combinations of the remaining three. Note that this identification is conditional on the chosen value $\delta_0 + \sigma_0^2$ in the auxiliary variable $\hat{r}_t$ that enters the regressors in (A.1). When iterating, this value is updated, as exactly the parameter combination $\delta + \sigma^2$ is one of the five that are conditionally identified. Ultimately, this identifies these five parameter functions. Instead of obtaining the five parameter functions, one can impose restrictions on ρ, δ, or σ^2 to identify all other parameters. Instead, without the need for such additional restrictions, it is possible to separate ρ, δ, and σ^2, and thus identify all six structural parameters, by exploiting the functional form of the error variances (the variances of (A.3a)-(A.3c)). Indeed, including the variance of the residual (A.3a) from the consumption equation as a separate moment along with the relations (A.2a)-(A.2c) clearly identifies σ^2 and thereby the full parameter vector ϕ, i.e., all six structural parameters.

Why should we rely on the first moment conditions and thus the regression coefficients, only, if they do not identify all structural parameters? In models with, say, stochastic volatility or more elaborate preference specification, the error term of the consumption equation becomes intractable (like the residual of the output equation). In such a case, the econometrician may exploit the martingale property only, without considering second moment conditions - namely, the form of the error variances and covariances. Because we want to keep our analysis applicable to such specifications, we focus on how to estimate the (identified) parameters from first moments in the main text, without going to higher moments. For comparison we show the results if we used the residual variance of the consumption and the interest rate equation in this web appendix.

In the given setup, with either five or six structural parameters thus identified, we extract estimates of them from the OLS, SUR, or FGLS-SUR-IV reduced-form parameter estimates using a minimum distance approach. We carry out minimum distance estimation based on either of three different unrestricted parameter sets ω_i, $i = 1, 2, 3$, from the reduced-form regressions: (1) the estimates of β in (A.2), i.e., the theoretical and empirical moments to match with respect to choice of ϕ are $\omega_1(\phi) = \beta(\phi)$ and $\hat{\omega}_1 = \hat{\beta}$ (this is the first moments or regression coefficients only case); (2) β along with the variance $\sigma^2\Delta$ of the consumption equation residual in (A.3a), so that $\omega_2(\phi) = (\omega_1(\phi)^\top, \sigma^2\Delta)^\top$ and $\hat{\omega}_2 = (\hat{\omega}_1^\top, \hat{\Sigma}_{CC})^\top$, with $\hat{\Sigma}_{CC}$ the upper left entry in the residual covariance matrix $\hat{\Sigma}$; (3) β along with the variances of the consumption and interest rate residuals (A.3a) and (A.3c), $\omega_3(\phi) = (\omega_2(\phi)^\top, \frac{1}{2}\eta^2(1 - e^{-2\kappa\Delta})/\kappa)^\top$ and $\hat{\omega}_3 = (\hat{\omega}_2^\top, \hat{\Sigma}_{rr})^\top$. In each of the three cases, we solve the problem

$$\hat{\phi} = \arg\min_{\phi} \left(\omega_i(\phi) - \hat{\omega}_i\right)^\top \hat{\Omega}_i^{-1} \left(\omega_i(\phi) - \hat{\omega}_i\right).$$

Here, the relevant metrics are given by the precisions of the reduced form estimates,

$$\hat{\Omega}_1^{-1} = \begin{pmatrix} \hat{\Sigma}^{CC} x_C^\top x_C & \hat{\Sigma}^{CY} x_C^\top x_Y & \hat{\Sigma}^{Cr} x_C^\top x_r \\ \hat{\Sigma}^{YC} x_Y^\top x_C & \hat{\Sigma}^{YY} x_Y^\top x_Y & \hat{\Sigma}^{Yr} x_Y^\top x_r \\ \hat{\Sigma}^{rC} x_r^\top x_C & \hat{\Sigma}^{rY} x_r^\top x_Y & \hat{\Sigma}^{rr} x_r^\top x_r \end{pmatrix},$$

$$\hat{\Omega}_2^{-1} = \begin{pmatrix} \hat{\Omega}_1^{-1} & 0_{6\times 1} \\ 0_{1\times 6} & \left(2\hat{\Sigma}_{CC}^2\right)^{-1} \end{pmatrix}, \qquad \hat{\Omega}_3 = \begin{pmatrix} \hat{\Omega}_2^{-1} & 0_{7\times 1} \\ 0_{1\times 7} & \left(2\hat{\Sigma}_{rr}^2\right)^{-1} \end{pmatrix},$$

with $\hat{\Sigma}^{ij}$ the (i, j)'th entry in $\hat{\Sigma}^{-1}$.

The indicated matrix $\hat{\Omega}_1^{-1}$ is for the case where the reduced form estimates $\hat{\beta}$ are obtained using SUR, i.e., $\hat{\Omega}_1 = \hat{V}_{SUR}$. If $\hat{\beta}$ is instead obtained by OLS as in (A.4), then the correct

$\hat{\Omega}_1 = \hat{V}_{OLS}$ is given in Appendix A.2. A naive OLS assessment of $\hat{\Omega}_1^{-1}$ would have zero off-diagonal blocks, and diagonal blocks $\hat{\Sigma}_{jj}^{-1} x_j^\top x_j$ in the minimum distance approach. With endogeneity correction, i.e., the reduced form estimates are obtained by FGLS-SUR-IV, again the minimum distance approach requires a variance-covariance matrix, and this has the same form as in the SUR case, but with the new $\hat{\Sigma}$ from the FGLS-SUR-IV approach and with fitted values for the relevant portions of x.

In case (1), using first moment conditions and thus β, only, to set up the minimum distance problem, estimators that are asymptotically equivalent to the resulting minimum distance estimators are alternatively obtained by restricted (nonlinear) regression, minimizing the OLS respectively the SUR objective function with respect to ϕ under the relevant restrictions (A.2a)-(A.2c) on β. In particular, the OLS objective is $\sum_{j=C,Y,r} \varepsilon_j^\top \varepsilon_j / \hat{\Sigma}_{jj}$ and the SUR objective $\sum_{t=1}^T \varepsilon_t^\top \hat{\Sigma}^{-1} \varepsilon_t$, where ε_j and ε_t are residual vectors of dimension T and 3, respectively, with elements $\varepsilon_{j,t}$. In cases (2) and (3), when estimated residual variances are used along with the relations (A.2a)-(A.2c) to identify structural parameters in the minimum distance case, then an asymptotically equivalent estimator may be obtained by iterating on structural parameters as they enter both ε_t and $\Sigma = \Sigma(\phi)$, used instead of $\hat{\Sigma}$ in the modified SUR objective function, say, $SUR^*(\phi) = \sum_{t=1}^T \varepsilon(\phi)_t^\top \Sigma(\phi)^{-1} \varepsilon(\phi)_t$, or, even better, $T \log |\Sigma(\phi)| + SUR^*(\phi)$. This use of (minus twice) the Gaussian log-likelihood function amounts to quasi maximum likelihood (QML) since clearly $\varepsilon_{Y,t}$ in (A.3b) is non-Gaussian.

A.2 The SUR estimator

The standard SUR assessment of the asymptotic variance-covariance matrix of $\hat{\beta}_{SUR}$ is $\hat{V}_{SUR} = (x^\top \hat{V}^{-1} x)^{-1}$. Note that the (i,j)'th block of the matrix being inverted is $\hat{\Sigma}^{ij} x_i^\top x_j$, with $\hat{\Sigma}^{ij}$ the (i,j)'th entry in $\hat{\Sigma}^{-1}$. Thus,

$$\hat{V}_{SUR} = \begin{pmatrix} \hat{\Sigma}^{CC} x_C^\top x_C & \hat{\Sigma}^{CY} x_C^\top x_Y & \hat{\Sigma}^{Cr} x_C^\top x_r \\ \hat{\Sigma}^{YC} x_Y^\top x_C & \hat{\Sigma}^{YY} x_Y^\top x_Y & \hat{\Sigma}^{Yr} x_Y^\top x_r \\ \hat{\Sigma}^{rC} x_r^\top x_C & \hat{\Sigma}^{rY} x_r^\top x_Y & \hat{\Sigma}^{rr} x_r^\top x_r \end{pmatrix}^{-1}.$$

If the covariances $\hat{\Sigma}_{ij}$ $(i \neq j)$ are zero, then the estimated asymptotic variance of $\hat{\beta}_j$ coincides with the OLS assessment $\hat{\Sigma}_{jj}(x_j^\top x_j)^{-1}$. More generally, the SUR approach suggests that the variance-covariance matrix $\hat{V}_{OLS}$ of the unrestricted OLS estimator from (A.4) has blocks estimated as $\hat{\Sigma}_{ij}(x_i^\top x_i)^{-1} x_i^\top x_j (x_j^\top x_j)^{-1}$, i.e., $\hat{V}_{OLS}$ equals

$$\begin{pmatrix} \hat{\Sigma}_{CC}(x_C^\top x_C)^{-1}(x_C^\top x_C)(x_C^\top x_C)^{-1} & \hat{\Sigma}_{CY}(x_C^\top x_C)^{-1}(x_C^\top x_Y)(x_Y^\top x_Y)^{-1} & \hat{\Sigma}_{Cr}(x_C^\top x_C)^{-1}(x_C^\top x_r)(x_r^\top x_r)^{-1} \\ \hat{\Sigma}_{YC}(x_Y^\top x_Y)^{-1}(x_Y^\top x_C)(x_C^\top x_C)^{-1} & \hat{\Sigma}_{YY}(x_Y^\top x_Y)^{-1}(x_Y^\top x_Y)(x_Y^\top x_Y)^{-1} & \hat{\Sigma}_{Yr}(x_Y^\top x_Y)^{-1}(x_Y^\top x_r)(x_r^\top x_r)^{-1} \\ \hat{\Sigma}_{rC}(x_r^\top x_r)^{-1}(x_r^\top x_C)(x_C^\top x_C)^{-1} & \hat{\Sigma}_{rY}(x_r^\top x_r)^{-1}(x_r^\top x_Y)(x_Y^\top x_Y)^{-1} & \hat{\Sigma}_{rr}(x_r^\top x_r)^{-1}(x_r^\top x_r)(x_r^\top x_r)^{-1} \end{pmatrix}^{-1}$$

and $\hat{V}_{OLS} \geq \hat{V}_{SUR}$ in the partial order of positive semi-definite matrices.

A.3 Transition probability matrix

This section derives the transition probability matrix for the continuous-time Markov chain of the regime-switching model (cf. Section 3.3.2).

Consider the following question that we use in our estimation approach: If the volatility is high at time $s \leq t$, then what is the probability that volatility is high at time t?

Following Ross (2014, p.371), let $P_{ij}(t) \equiv P(\eta_t = \eta_j \mid \eta_s = \eta_i)$ for $s \leq t$ denote the probability that a process presently in state i will be in state j at time t, and ϕ_{ij} the instantaneous transition rates, when in state i, at which the process makes a transition into state j. We shall derive the desired probability, namely $P_{hh}(t)$ by solving

$$\begin{aligned} \dot{P}_{hh}(t) &= \phi_{hl}\left[P_{lh}(t) - P_{hh}(t)\right], \\ \dot{P}_{lh}(t) &= \phi_{lh}\left[P_{hh}(t) - P_{lh}(t)\right], \end{aligned}$$

$\dot{P}_{ij}(t) \equiv \lim_{h\to 0}[P_{ij}(t+h) - P_{ij}(t)]/h$ for all $i, j \in \Theta$ with initial conditions $P_{hh}(s) = 1$ and $P_{lh}(s) = 0$. The solution to this system of ODEs is given by:

$$P_{hh}(t) = \frac{\phi_{lh}}{\phi_{hl}+\phi_{lh}} + \frac{\phi_{hl}}{\phi_{hl}+\phi_{lh}} e^{-(\phi_{hl}+\phi_{lh})(t-s)}, \tag{A.6}$$

$$P_{lh}(t) = \frac{\phi_{lh}}{\phi_{hl}+\phi_{lh}} - \frac{\phi_{lh}}{\phi_{hl}+\phi_{lh}} e^{-(\phi_{hl}+\phi_{lh})(t-s)}. \tag{A.7}$$

Hence, the transition probability matrix of the continuous-time Markov chain for $s \leq t$ is

$$P(t) = \begin{bmatrix} P_{ll}(t) & P_{lh}(t) \\ P_{hl}(t) & P_{hh}(t) \end{bmatrix}, \tag{A.8}$$

in which $P_{ll}(t) = 1 - P_{lh}(t)$ and $P_{hl}(t) = 1 - P_{hh}(t)$.

If we let $P_h(s)$ denote the unconditional probability of being in state θ_h at time s, the unconditional probability of being in the same state at time $t > s$ is then

$$P_h(t) = P_h(s)P_{hh}(t) + (1 - P_h(s))\, P_{lh}(t).$$

In the limit as $t \to \infty$ the unconditional probability of being in the high regime is

$$\lim_{t\to\infty} P_h(t) = \frac{\phi_{lh}}{\phi_{hl}+\phi_{lh}}.$$

A similar procedure yields the unconditional probability of being in the low regime as

$$\lim_{t\to\infty} P_l(t) = \frac{\phi_{hl}}{\phi_{hl}+\phi_{lh}}.$$

Conversely, from (A.6) and (A.7) we can back out the instantaneous transition rates of the Poisson processes, ϕ_{hl} and ϕ_{lh}, from any given transition probability matrix.

B Comparison to the discrete-time model

We now develop the model in discrete-time formulation in order to compare both approaches. Before we start it is important to note that all state variables are directly comparable, whereas the flow variables are expressed as periodic rates (instead of instantaneous rates).

B.1 The model

Production possibilities. For the ease of readability, we present the full model below. The production function is a constant returns to scale technology

$$Y_t = A_t F(K_t, L), \tag{B.1}$$

where K_t is the (predetermined) aggregate capital stock, L is the constant population size, and A_t is total factor productivity, which follows an autoregressive process

$$A_{t+1} - A_t = \tilde{\mu}(A_t) + \tilde{\eta}(A_t)\epsilon_{A,t+1}, \quad \epsilon_A \sim \mathcal{N}(0,1), \tag{B.2}$$

with $\mu(A_t)$ and $\eta(A_t)$ generic drift and volatility functions.[2] The capital stock increases if gross investment I_t exceeds capital depreciation,

$$K_{t+1} - K_t = I_t - \tilde{\delta}K_t + \tilde{\sigma}K_t\epsilon_{K,t+1}, \quad \epsilon_K \sim \mathcal{N}(0,1), \tag{B.3}$$

where $\tilde{\delta}$ is a deterministic rate of depreciation and $\tilde{\sigma}$ determines the variance of the stochastic depreciation.[3] Similar to the continuous-time version, the stochastic depreciation does depend on the level of the predetermined capital stock.

Equilibrium properties. In equilibrium, factors of production are rewarded with marginal products $\tilde{r}_t = Y_K$ and $\tilde{w}_t = Y_L$, subscripts K and L indicating derivatives, and the goods market clears, $Y_t = C_t + I_t$. Although there is no stochastic calculus for discrete-time models, we may express the evolution of equilibrium output in this economy as

$$Y_{t+1} = \left(A_t + \tilde{\mu}(A_t) + \tilde{\eta}(A_t)\epsilon_{A,t+1}\right) F(K_t + I_t - \tilde{\delta}K_t + \tilde{\sigma}K_t\epsilon_{K,t+1}, L). \tag{B.4}$$

Preferences. Consider an economy with a single consumer, interpreted as a representative "stand-in" for a large number of identical consumers. The consumer maximizes expected additively separable discounted life-time utility given by

$$U_0 \equiv E_0 \sum_{t=0}^{\infty} \tilde{\beta}^t u(C_t, A_t)dt, \quad u_C > 0, \; u_{CC} < 0, \tag{B.5}$$

[2]We assume that $E(A_t) = A \in \mathbb{R}_+$ exists, and that the sum describing life-time utility in (B.5) below is bounded, so that the value function is well-defined.

[3]It is insightful to relate the two shocks in the system to the continuous-time counterpart by looking at the Euler approximation $\epsilon_{A,t+1} \equiv B_{t+1} - B_t \sim \mathcal{N}(0,1)$ and $\epsilon_{K,t+1} \equiv Z_{t+1} - Z_t \sim \mathcal{N}(0,1)$.

subject to

$$K_{t+1} - K_t = (\tilde{r}_t - \tilde{\delta})K_t + \tilde{w}_t L - C_t + \tilde{\sigma} K_t \epsilon_{K,t+1}, \tag{B.6}$$

where $\tilde{\beta}$ is the subjective discount factor, $\tilde{r}_t$ is the rental rate of capital, and $\tilde{w}_t$ is the labor wage rate. The paths of factor rewards are taken as given by the representative consumer.

B.2 The Euler equation

The relevant state variables are capital and technology, (K_t, A_t). For given initial states, the value of the optimal program is

$$V(K_0, A_0) = \max_{\{C_t\}_{t=0}^{\infty}} U_0 \quad s.t. \quad (B.6) \quad \text{and} \quad (B.2), \tag{B.7}$$

i.e., the present value of expected utility along the optimal program. As a necessary condition for optimality, Bellman's principle gives at time s

$$V(K_s, A_s) = \max_{C_s} \left\{ u(C_s, A_s) + \tilde{\beta} E_s \left[V(K_{s+1}, A_{s+1}) \right] \right\}. \tag{B.8}$$

Hence, the first-order condition for the problem is

$$u_C(C_t, A_t) = \tilde{\beta} E_t \left[V_K(K_{t+1}, A_{t+1}) \right], \tag{B.9}$$

for any $t \in [0, \infty)$, and this allows us to write consumption as a function of the state variables, $C_t = C(K_t, A_t)$. Hence, the discrete-time model requires evaluating an integral (integrating out expectations) to obtain the optimal consumption function. The reason is that the Hamilton-Jacobi-Bellman (HJB) equation in the discrete-time model (B.8) requires to solve a stochastic difference equation in contrast to a deterministic differential equation.

Using the concentrated Bellman equation,

$$V(K_t, A_t) = u(C(K_t, A_t)) + \tilde{\beta} E_t V(K_{t+1}, A_{t+1})$$

we obtain

$$\begin{aligned} V_K(K_t, A_t) &= \tilde{\beta} E_t \left[V_K(K_{t+1}, A_{t+1})(1 - \tilde{\delta} + \tilde{r}_t + \tilde{\sigma} \epsilon_{K,t+1}) \right] \\ &= (1 - \tilde{\delta} + \tilde{r}_t) u_C(C_t, A_t) + \tilde{\beta} E_t \left[V_K(K_{t+1}, A_{t+1}) \tilde{\sigma} \epsilon_{K,t+1} \right]. \end{aligned}$$

Note that the second term is zero in equilibrium, because from the first-order condition

$$\begin{aligned} u_C(C_t, A_t) \tilde{\sigma} \epsilon_{K,t+1} &= \tilde{\beta} E_t \left[V_K(K_{t+1}, A_{t+1}) \right] \tilde{\sigma} \epsilon_{K,t+1} \\ \Leftrightarrow \ E_t \left[u_C(C_t, A_t) \tilde{\sigma} \epsilon_{K,t+1} \right] &= \tilde{\beta} E_t \left[E_t \left[V_K(K_{t+1}, A_{t+1}) \right] \tilde{\sigma} \epsilon_{K,t+1} \right] \\ \Leftrightarrow \ u_C(C_t, A_t) \tilde{\sigma} E_t \left[\epsilon_{K,t+1} \right] &= \tilde{\beta} E_t \left[V_K(K_{t+1}, A_{t+1}) \tilde{\sigma} \epsilon_{K,t+1} \right] \\ \Leftrightarrow \ 0 &= \tilde{\beta} E_t \left[V_K(K_{t+1}, A_{t+1}) \tilde{\sigma} \epsilon_{K,t+1} \right] \end{aligned}$$

Hence,

$$\begin{aligned} V_K(K_t, A_t) &= \tilde{\beta} E_t \left[V_K(K_{t+1}, A_{t+1})(1 - \tilde{\delta} + \tilde{r}_t) \right] \\ &= (1 - \tilde{\delta} + \tilde{r}_t) u_C(C_t, A_t). \end{aligned}$$

Leading the expression one period ahead and applying expectations yields

$$E_t \left[V_K(K_{t+1}, A_{t+1}) \right] = E_t \left[(1 - \tilde{\delta} + \tilde{r}_{t+1}) u_C(C_{t+1}, A_{t+1}) \right].$$

Inserting back into the first-order condition (B.9) we arrive at the Euler equation

$$u_C(C_t, A_t) = \tilde{\beta} E_t \left[(1 - \tilde{\delta} + \tilde{r}_{t+1}) u_C(C_{t+1}, A_{t+1}) \right], \tag{B.10}$$

In the following, we restrict attention to the case $u(C_t, A_t) = u(C_t)$.

B.3 Equilibrium dynamics

Our equilibrium dynamics of the economy can be summarized as

$$\begin{aligned} u'(C_t) &= \tilde{\beta} E_t \left[(1 - \tilde{\delta} + \tilde{r}_{t+1}) u'(C_{t+1}) \right] && \text{(B.11a)} \\ Y_{t+1} &= (A_t + \tilde{\mu}(A_t) + \tilde{\eta}(A_t)\epsilon_{A,t+1}) \, F(K_t + I_t - \tilde{\delta} K_t + \tilde{\sigma} K_t \epsilon_{K,t+1}, L) && \text{(B.11b)} \\ K_{t+1} &= (1 + \tilde{r}_t - \tilde{\delta}) K_t + \tilde{w}_t L - C_t + \tilde{\sigma} K_t \epsilon_{K,t+1} && \text{(B.11c)} \\ A_{t+1} &= A_t + \tilde{\mu}(A_t) + \tilde{\eta}(A_t)\epsilon_{A,t+1} && \text{(B.11d)} \end{aligned}$$

Provided that variables C_t, Y_t, K_t and also A_t are observed, the econometrician needs to consider the system (B.11) for statistical inference on the deep parameters.

For comparison, the equilibrium dynamics the corresponding continuous-time economy of the model used in the main text can be summarized as

$$\begin{aligned} dC_t &= \frac{u'(C_t)}{u''(C_t)}(\rho - (r_t - \delta))dt - \sigma^2 C_K K_t dt - \tfrac{1}{2}(C_A^2 \eta(A_t)^2 + C_K^2 \sigma^2 K_t^2) \frac{u'''(C_t)}{u''(C_t)} dt \\ &\quad + C_A \eta(A_t) dB_t + C_K \sigma K_t dZ_t && \text{(B.12a)} \\ dY_t &= Y_A dA_t + Y_K dK_t + \tfrac{1}{2} Y_{KK} \sigma^2 K_t^2 dt \\ &= (\mu(A_t) Y_A + (I_t - \delta K_t) Y_K + \tfrac{1}{2} Y_{KK} \sigma^2 K_t^2) dt + Y_A \eta(A_t) dB_t + \sigma Y_K K_t dZ_t && \text{(B.12b)} \\ dK_t &= (I_t - \delta K_t) dt + \sigma K_t dZ_t && \text{(B.12c)} \\ dA_t &= \mu(A_t) dt + \eta(A_t) dB_t && \text{(B.12d)} \end{aligned}$$

Provided that C_t, Y_t, K_t and also A_t are observed, the econometrician needs to consider the system (B.12) for statistical inference on the deep parameters.

In what follows, we assume that the capital stock K_t and A_t are latent variables, but we can obtain them from financial market data.

B.4 The AK-Vasicek model with logarithmic preferences

Consider an AK economy, $Y_t = A_t K_t$, which implies $\tilde{r}_t = A_t$ and $K_t = Y_t/\tilde{r}_t$, and assume that the consumer has logarithmic preferences, system (B.11) reduces to,

$$\begin{aligned} C_t^{-1} &= \tilde{\beta} E_t \left[(1 - \tilde{\delta} + \tilde{r}_{t+1}) C_{t+1}^{-1} \right] && \text{(B.13a)} \\ Y_{t+1} &= (\tilde{r}_t + \tilde{\mu}(\tilde{r}_t) + \tilde{\eta}(\tilde{r}_t)\epsilon_{A,t+1}) \left((1 + \tilde{r}_t - \tilde{\delta})(Y_t/\tilde{r}_t) - C_t + \tilde{\sigma} Y_t/\tilde{r}_t \epsilon_{K,t+1} \right) && \text{(B.13b)} \\ \tilde{r}_{t+1} &= \tilde{r}_t + \tilde{\mu}(\tilde{r}_t) + \tilde{\eta}(\tilde{r}_t)\epsilon_{A,t+1} && \text{(B.13c)} \end{aligned}$$

whereas system (B.12) reduces to

$$\begin{aligned} dC_t &= (r_t - \delta - \rho) C_t dt - \sigma^2 C_K K_t dt - (C_A^2 \eta(A_t)^2 + C_K^2 \sigma^2 K_t^2)/C_t dt \\ &\quad + C_A \eta(A_t) dB_t + C_K \sigma Y_t / r_t dZ_t && \text{(B.14a)} \\ dY_t &= (\mu(r_t) Y_t / r_t + (r_t - \delta) Y_t - r_t C_t) dt + \eta(r_t) Y_t / r_t dB_t + \sigma Y_t dZ_t && \text{(B.14b)} \\ dr_t &= \mu(r_t) dt + \eta(r_t) dB_t && \text{(B.14c)} \end{aligned}$$

Both systems give the model in terms of observables (macro and financial market data).

The Vasicek (1977) mean-reverting specification for the rental rate of physical capital is $\mu(r_t) = \kappa(\gamma - r_t)$ and $\eta(r_t) = \eta$, where $\kappa > 0$ is the speed and γ the target rate of mean reversion, and η the constant volatility. The corresponding Vasicek mean-reversion model at quarterly frequency reads $\tilde{\mu}(\tilde{r}_t) = \tilde{\kappa}(\tilde{\gamma} - \tilde{r}_t)$ and $\tilde{\eta}(\tilde{r}_t) = \tilde{\eta}$ where we define

$$\tilde{\gamma} \equiv \Delta\gamma, \quad \tilde{\kappa} \equiv 1 - e^{-\Delta\kappa}, \quad \tilde{\eta} \equiv \Delta\eta \sqrt{(1 - e^{-2\kappa\Delta})/(2\kappa)} \tag{B.15}$$

In this case, the equilibrium dynamics are

$$\begin{aligned} C_t^{-1} &= \tilde{\beta} E_t \left[(1 - \tilde{\delta} + \tilde{r}_{t+1}) C_{t+1}^{-1} \right] && \text{(B.16a)} \\ Y_{t+1} &= Y_t + (\tilde{r}_t - \tilde{\delta}) Y_t - \tilde{r}_t C_t + \tilde{\kappa}(\tilde{\gamma} - \tilde{r}_t) Y_t / \tilde{r}_t + \tilde{\eta} Y_t / \tilde{r}_t \epsilon_{A,t+1} + \tilde{\sigma} Y_t \epsilon_{K,t+1} \\ &\quad + ((\tilde{r}_t - \tilde{\delta}) Y_t / \tilde{r}_t - C_t + \tilde{\sigma} Y_t / \tilde{r}_t \epsilon_{K,t+1}) (\tilde{\kappa}(\tilde{\gamma} - \tilde{r}_t) + \tilde{\eta}\epsilon_{A,t+1}) && \text{(B.16b)} \\ \tilde{r}_{t+1} &= \tilde{r}_t + \tilde{\kappa}(\tilde{\gamma} - \tilde{r}_t) + \tilde{\eta}\epsilon_{A,t+1} && \text{(B.16c)} \end{aligned}$$

whereas system (B.14) reads

$$\begin{aligned} dC_t &= (r_t - \delta - \rho) C_t dt - \sigma^2 C_K K_t dt - (C_A^2 \eta^2 + C_K^2 \sigma^2 K_t^2)/C_t dt \\ &\quad + C_A \eta dB_t + C_K \sigma Y_t / r_t dZ_t && \text{(B.17a)} \\ dY_t &= ((\kappa\gamma / r_t - \kappa + r_t - \delta) Y_t - r_t C_t) dt + \eta Y_t / r_t dB_t + \sigma Y_t dZ_t && \text{(B.17b)} \\ dr_t &= \kappa(\gamma - r_t) dt + \eta dB_t && \text{(B.17c)} \end{aligned}$$

Before we estimate the structural parameters, we need to solve the two models. This is complicated by the fact that both models are highly nonlinear. Note that the AK-Vasicek model with logarithmic preferences in continuous time has an explicit analytical solution of the nonlinear system, whereas the discrete-time analogue can only be solved numerically. We follow a log-linear approximation of the discrete-time first-order conditions below and use log-linear representation of the equilibrium dynamics for estimation.

One simple way of proceeding with the continuous-time system in order to match it to the discrete-time nature of the data is to use an Euler scheme (as in Wang, Phillips, and Yu, 2011) to discretize the system (B.14) for small time intervals (no approximation error in the limit). This scheme has the nice feature that the discrete-time econometric toolbox (i.e., either linear or nonlinear estimation methods following An and Schorfheide, 2007; Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007; Fernández-Villaverde and Rubio-Ramírez, 2007) can be applied and thus seems quite attractive. As explained in the main text, we do not follow this route. Instead we proceed by integrating the system of equations and/or use closed-form solutions, for example for the interest rate Vasicek specification. This allows us to easily handle different frequencies for the estimation of structural parameters.

B.5 Log-linear approximation, discrete-time AK-Vasicek model

There are many ways to solve the discrete-time model numerically. The best practice is to solve the model through a log-linear approximation to the set of first-order conditions. For this we define auxiliary variables (which turn out to be stationary),

$$\hat{C}_t \equiv \frac{C_t}{K_t}, \quad 1 + \nu_{t+1} \equiv \frac{K_{t+1}}{K_t}$$

and which can be used to transform the Euler equation (B.16a) into

$$\begin{aligned} 1 &= \tilde{\beta} E_t \left[(1 - \tilde{\delta} + \tilde{r}_{t+1}) \frac{C_t}{C_{t+1}} \frac{K_{t+1}}{K_{t+1}} \frac{K_t}{K_t} \right] \\ &= \tilde{\beta} E_t \left[\frac{\hat{C}_t}{\hat{C}_{t+1}} \frac{1 - \tilde{\delta} + \tilde{r}_{t+1}}{1 + \nu_{t+1}} \right] \end{aligned} \tag{B.18}$$

We may write the aggregate resource constraint (B.16b) as

$$\begin{aligned} Y_{t+1} &= \tilde{r}_{t+1}(K_t + Y_t - C_t - \tilde{\delta} K_t + \tilde{\sigma} K_t \epsilon_{K,t+1}) \\ \Leftrightarrow \frac{K_{t+1}}{K_t} &= 1 + \frac{Y_t}{K_t} - \frac{C_t}{K_t} - \tilde{\delta} + \tilde{\sigma} \epsilon_{K,t+1} \end{aligned}$$

or

$$\nu_{t+1} = \tilde{r}_t - \hat{C}_t - \tilde{\delta} + \tilde{\sigma} \epsilon_{K,t+1} \tag{B.19}$$

As a reference level, with $\tilde{r}_t \equiv \tilde{\gamma}$ for all t, the non-stochastic steady-state value is given from the Euler equation, which implies steady-state value for the consumption-capital ratio

$$1 + \nu = \tilde{\beta}(1 - \tilde{\delta} + \tilde{\gamma}) \quad \Rightarrow \quad \hat{C} = (1 - \tilde{\beta})(1 - \tilde{\delta} + \tilde{\gamma}) \tag{B.20}$$

First, we rewrite (B.18) as

$$\frac{1}{\tilde{\beta}} = \frac{\hat{C}_t}{\hat{C}_{t+1}} \frac{1-\tilde{\delta}+\tilde{r}_{t+1}}{1-\tilde{\delta}+\tilde{r}_t-\hat{C}_t+\tilde{\sigma}\epsilon_{K,t+1}} - v_{t+1} \;\equiv\; G(\ln \hat{C}_{t+1}, \ln \hat{C}_t, \tilde{r}_{t+1}, \tilde{r}_t, \epsilon_{K,t+1}, v_{t+1})$$

in which we defined the expectations error

$$v_{t+1} \equiv \frac{\hat{C}_t}{\hat{C}_{t+1}} \frac{1-\tilde{\delta}+\tilde{r}_{t+1}}{1+\nu_{t+1}} - E_t\left[\frac{\hat{C}_t}{\hat{C}_{t+1}} \frac{1-\tilde{\delta}+\tilde{r}_{t+1}}{1+\nu_{t+1}}\right], \quad E_t(v_{t+1}=0)$$

Second, we log-linearize the equation about the non-stochastic steady-state values

$$\begin{aligned} 0 \;\simeq\;& -\frac{1}{\tilde{\beta}}(\ln \hat{C}_{t+1} - \ln \hat{C}) + \frac{1}{\tilde{\beta}^2}(\ln \hat{C}_t - \ln \hat{C}) + \frac{1-\tilde{\beta}}{\tilde{\beta}\hat{C}}(\tilde{r}_{t+1}-\tilde{\gamma}) - \frac{1-\tilde{\beta}}{\tilde{\beta}^2\hat{C}}(\tilde{r}_t-\tilde{\gamma}) \\ & -\frac{\tilde{\sigma}(1-\tilde{\beta})}{\tilde{\beta}^2\hat{C}}\epsilon_{K,t+1} - v_{t+1} \end{aligned}$$

or

$$\begin{aligned} \ln \hat{C}_{t+1} - \ln \hat{C} - \frac{1-\tilde{\beta}}{\hat{C}}(\tilde{r}_{t+1}-\tilde{\gamma}) \;\simeq\;& \frac{1}{\tilde{\beta}}(\ln \hat{C}_t - \ln \hat{C}) - \frac{1-\tilde{\beta}}{\tilde{\beta}\hat{C}}(\tilde{r}_t-\tilde{\gamma}) \\ & -\frac{\tilde{\sigma}(1-\tilde{\beta})}{\tilde{\beta}\hat{C}}\epsilon_{K,t+1} - v_{t+1} \end{aligned}$$

where we used

$$\begin{aligned} \left.\frac{\partial G}{\partial \ln \hat{C}_{t+1}}\right|_{ss} &= \left.-\frac{\hat{C}_t}{\hat{C}_{t+1}} \frac{1-\tilde{\delta}+\tilde{r}_{t+1}}{1-\tilde{\delta}+\tilde{r}_t-\hat{C}_t+\tilde{\sigma}\epsilon_{K,t+1}}\right|_{ss} = -\frac{1-\tilde{\delta}+\tilde{\gamma}}{1-\tilde{\delta}+\tilde{\gamma}-\hat{C}} = -\frac{1}{\tilde{\beta}}, \\ \left.\frac{\partial G}{\partial \ln \hat{C}_t}\right|_{ss} &= \left.\frac{\hat{C}_t}{\hat{C}_{t+1}} \frac{1-\tilde{\delta}+\tilde{r}_{t+1}}{1-\tilde{\delta}+\tilde{r}_t-\hat{C}_t+\tilde{\sigma}\epsilon_{K,t+1}} - \frac{\hat{C}_t}{\hat{C}_{t+1}} \frac{1-\tilde{\delta}+\tilde{r}_{t+1}}{(1-\tilde{\delta}+\tilde{r}_t-\hat{C}_t+\tilde{\sigma}\epsilon_{K,t+1})^2}(-\hat{C}_t)\right|_{ss} \\ &= \frac{1-\tilde{\delta}+\tilde{\gamma}}{1-\tilde{\delta}+\tilde{\gamma}-\hat{C}} + \hat{C}\frac{1-\tilde{\delta}+\tilde{\gamma}}{(1-\tilde{\delta}+\tilde{\gamma}-\hat{C})^2} = \frac{1}{\tilde{\beta}^2} \\ \left.\frac{\partial G}{\partial r_{t+1}}\right|_{ss} &= \left.\frac{\hat{C}_t}{\hat{C}_{t+1}} \frac{1}{1-\tilde{\delta}+\tilde{r}_t-\hat{C}_t+\tilde{\sigma}\epsilon_{K,t+1}}\right|_{ss} = \frac{1}{1-\tilde{\delta}+\tilde{\gamma}-\hat{C}} = \frac{1-\tilde{\beta}}{\tilde{\beta}\hat{C}} \\ \left.\frac{\partial G}{\partial r_t}\right|_{ss} &= \left.-\frac{\hat{C}_t}{\hat{C}_{t+1}} \frac{1-\tilde{\delta}+\tilde{r}_{t+1}}{(1-\tilde{\delta}+\tilde{r}_t-\hat{C}_t+\tilde{\sigma}\epsilon_{K,t+1})^2}\right|_{ss} = -\frac{1-\tilde{\delta}+\tilde{\gamma}}{(1-\tilde{\delta}+\tilde{\gamma}-\hat{C})^2} = -\frac{1-\tilde{\beta}}{\tilde{\beta}^2\hat{C}} \\ \left.\frac{\partial G}{\partial \epsilon_{K,t+1}}\right|_{ss} &= \left.-\frac{\hat{C}_t}{\hat{C}_{t+1}} \frac{1-\tilde{\delta}+\tilde{r}_{t+1}}{(1-\tilde{\delta}+\tilde{r}_t-\hat{C}_t+\tilde{\sigma}\epsilon_{K,t+1})^2}\tilde{\sigma}\right|_{ss} = -\frac{\tilde{\sigma}(1-\tilde{\beta})}{\tilde{\beta}^2\hat{C}} \end{aligned}$$

so that we get the matrix system

$$\begin{aligned} \begin{pmatrix} 1 & -(1-\tilde{\beta})/\hat{C} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \ln \hat{C}_{t+1} - \ln \hat{C} \\ \tilde{r}_{t+1}-\tilde{\gamma} \end{pmatrix} \;=\;& \begin{pmatrix} 1/\tilde{\beta} & -(1-\tilde{\beta})/(\tilde{\beta}\hat{C}) \\ 0 & 1-\tilde{\kappa} \end{pmatrix} \begin{pmatrix} \ln \hat{C}_t - \ln \hat{C} \\ \tilde{r}_t - \tilde{\gamma} \end{pmatrix} \\ & + \begin{pmatrix} -\tilde{\epsilon}_{K,t+1} - v_{t+1} \\ \tilde{\eta}\epsilon_{A,t+1} \end{pmatrix} \end{aligned}$$

where $\tilde{\epsilon}_{K,t+1} \equiv (\tilde{\sigma}(1-\tilde{\beta})/(\tilde{\beta}\hat{C}))\epsilon_{K,t+1}$. The matrix equation is of the form

$$\Phi_0 z_{t+1} = \Phi_1 z_t + \xi_{t+1} \quad \Rightarrow \quad z_{t+1} = (\Phi_0^{-1}\Phi_1) z_t + \Phi_0^{-1}\xi_{t+1}$$

where z_t is the vector of variables $z_t = (\ln \hat{C}_t - \ln \hat{C}, \tilde{r}_t - \tilde{\gamma})^\top$, and Φ_0 and Φ_1 are the coefficient matrices containing the structural parameters. Note that $\Phi_0^{-1}\Phi_1$ can be found as

$$\Phi_0^{-1}\Phi_1 = \begin{pmatrix} 1 & (1-\tilde{\beta})/\hat{C} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\tilde{\beta} & -(1-\tilde{\beta})/(\tilde{\beta}\hat{C}) \\ 0 & 1-\tilde{\kappa} \end{pmatrix} = \begin{pmatrix} 1/\tilde{\beta} & (1-\tilde{\beta})(1-\tilde{\kappa}-1/\tilde{\beta})/\hat{C} \\ 0 & 1-\tilde{\kappa} \end{pmatrix}$$

and the eigenvalues are obtained from the characteristic equation $|\Phi_0^{-1}\Phi_1 - \lambda I_2| = 0$ or

$$\begin{vmatrix} 1/\tilde{\beta} - \lambda & (1-\tilde{\beta})(1-\tilde{\kappa}-1/\tilde{\beta})/\hat{C} \\ 0 & 1-\tilde{\kappa}-\lambda \end{vmatrix} = (1/\tilde{\beta} - \lambda)(1-\tilde{\kappa}-\lambda) = 0$$

which yields $\lambda_1 = 1/\tilde{\beta}$ and $\lambda_2 = 1-\tilde{\kappa}$. While the latter is positive and less than 1, the first eigenvalue is greater than 1, that is, the economy will have a saddle path property, with a single trajectory leading to the unique steady state of the system.

Hence, we obtain the linear solution to the homogeneous matrix equation

$$z_t = \begin{pmatrix} \ln \hat{C}_t - \ln \hat{C} \\ \tilde{r}_t - \tilde{\gamma} \end{pmatrix} = \mathbb{C}_1 (1/\tilde{\beta})^t \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mathbb{C}_2 (1-\tilde{\kappa})^t \begin{pmatrix} (1-\tilde{\beta})/\hat{C} \\ 1 \end{pmatrix}$$

Because we need to focus on the stable path, the stability condition requires $\mathbb{C}_1 = 0$ and from the solution of the Vasicek specification we get $\mathbb{C}_2 = \tilde{r}_0 - \tilde{\gamma}$. Hence, we find that

$$\begin{aligned} \ln \hat{C}_t - \ln \hat{C} &= \frac{1-\tilde{\beta}}{\hat{C}} (\tilde{r}_t - \tilde{\gamma}) \\ \Leftrightarrow \ \ln(C_t/K_t) &= \ln \hat{C} + \frac{1}{1-\tilde{\delta}+\tilde{\gamma}} (\tilde{r}_t - \tilde{\gamma}) \end{aligned} \tag{B.21}$$

Given any value of r_0 and initial value K_0, we obtain the optimal level of consumption C_0, the next periods capital stock K_{t+1} is obtained from (B.19).[4] Using this solution, we get

$$\begin{aligned} \ln(C_{t+1}/C_t) - \ln(K_{t+1}/K_t) &= \frac{1}{1-\tilde{\delta}+\tilde{\gamma}} (\tilde{r}_{t+1} - \tilde{r}_t) \\ &= \frac{-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}} (\tilde{r}_t - \tilde{\gamma}) + \frac{1}{1-\tilde{\delta}+\tilde{\gamma}} \tilde{\eta}\epsilon_{A,t+1} \end{aligned}$$

Using a log-linear approximation of (B.19) and insert the solution such that

$$\begin{aligned} (1+\nu)(\ln(K_{t+1}/K_t) - \ln(1+\nu)) &\simeq \tilde{r}_t - \tilde{\gamma} - \hat{C}(\ln \hat{C}_t - \ln \hat{C}) + \tilde{\sigma}\epsilon_{K,t+1} \\ &= \tilde{r}_t - \tilde{\gamma} - (1-\tilde{\beta})(\tilde{r}_t - \tilde{\gamma}) + \tilde{\sigma}\epsilon_{K,t+1} \\ &= \tilde{\beta}(\tilde{r}_t - \tilde{\gamma}) + \tilde{\sigma}\epsilon_{K,t+1} \end{aligned}$$

[4] Note that solving a linear (instead a log-linear) approximation would imply $C_t/K_t = (1-\tilde{\beta})(1-\tilde{\delta}+\tilde{r}_t)$, which could be then log-linearized to arrive at the same result $\ln(C_t/K_t) = \ln \hat{C} + (\tilde{r}_t - \tilde{\gamma})/(1-\tilde{\delta}+\tilde{\gamma})$.

or

$$\ln(K_{t+1}/K_t) = \ln(1+\nu) + \frac{1}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t - \tilde{\gamma}) + \frac{\tilde{\sigma}}{1+\nu}\epsilon_{K,t+1} \tag{B.22}$$

yields

$$\ln(C_{t+1}/C_t) = \ln(1+\nu) + \frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t - \tilde{\gamma}) + \frac{\tilde{\sigma}}{1+\nu}\epsilon_{K,t+1} + \frac{\tilde{\eta}}{1-\tilde{\delta}+\tilde{\gamma}}\epsilon_{A,t+1} \tag{B.23}$$

such that the expectations error implied by our solution is (as from the Euler equation)

$$v_{t+1} \equiv -\frac{\tilde{\sigma}}{1+\nu}\epsilon_{K,t+1} \tag{B.24}$$

which implies that $E_t(v_{t+1}) = 0$ and $Var_t(v_{t+1}) = \tilde{\sigma}^2/(1+\nu)^2$.

We may use (B.22) together with a log-linear approximation of $Y_t = A_t K_t$ with $A_t = \tilde{r}_t$ or

$$\ln Y_t - \ln K_t \simeq \tilde{\gamma} + \frac{\tilde{r}_t - \tilde{\gamma}}{\tilde{\gamma}}$$

(such that the interest rate does not appear as logarithmic function) to obtain

$$\begin{aligned}
\ln(Y_{t+1}/Y_t) &= \ln(1+\nu) + \frac{\tilde{r}_{t+1} - \tilde{r}_t}{\tilde{\gamma}} + \frac{1}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t - \tilde{\gamma}) + \frac{\tilde{\sigma}}{1+\nu}\epsilon_{K,t+1} \\
&= \ln(1+\nu) + -\frac{\tilde{\kappa}}{\tilde{\gamma}}(\tilde{r}_t - \tilde{\gamma}) + \frac{1}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t - \tilde{\gamma}) + \frac{\tilde{\sigma}}{1+\nu}\epsilon_{K,t+1} + \frac{\tilde{\eta}}{\tilde{\gamma}}\epsilon_{A,t+1} \\
&= \ln(1+\nu) + \frac{\tilde{\gamma} - \tilde{\kappa}(1-\tilde{\delta}+\tilde{\gamma})}{1-\tilde{\delta}+\tilde{\gamma}}\frac{\tilde{r}_t - \tilde{\gamma}}{\tilde{\gamma}} + \frac{\tilde{\sigma}}{1+\nu}\epsilon_{K,t+1} + \frac{\tilde{\eta}}{\tilde{\gamma}}\epsilon_{A,t+1}
\end{aligned}$$

Summarizing, and using $\ln(1+\nu) = \ln\tilde{\beta} + \ln(1-\tilde{\delta}+\tilde{\gamma}) \approx \ln\tilde{\beta} - \tilde{\delta} + \tilde{\gamma}$ yields

$$\begin{aligned}
\ln(C_{t+1}/C_t) &= \ln\tilde{\beta} - \tilde{\delta} + \tilde{\gamma} + \frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t - \tilde{\gamma}) + \frac{\tilde{\sigma}\epsilon_{K,t+1}}{\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})} + \frac{\tilde{\eta}\epsilon_{A,t+1}}{1-\tilde{\delta}+\tilde{\gamma}} && \text{(B.25a)} \\
\ln(Y_{t+1}/Y_t) &= \ln\tilde{\beta} - \tilde{\delta} + \tilde{\gamma} + \frac{\tilde{\gamma} - \tilde{\kappa}(1-\tilde{\delta}+\tilde{\gamma})}{1-\tilde{\delta}+\tilde{\gamma}}\frac{\tilde{r}_t - \tilde{\gamma}}{\tilde{\gamma}} + \frac{\tilde{\sigma}\epsilon_{K,t+1}}{\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})} + \frac{\tilde{\eta}\epsilon_{A,t+1}}{\tilde{\gamma}} && \text{(B.25b)} \\
\tilde{r}_{t+1} &= \tilde{r}_t + \tilde{\kappa}(\tilde{\gamma} - \tilde{r}_t) + \tilde{\eta}\epsilon_{A,t+1} && \text{(B.25c)}
\end{aligned}$$

In this AK-Vasicek model, the relation between the one-period risk-free rate and the rental rate of capital is given by (see Section B.10)

$$(\tilde{r}_t^f - \tilde{\gamma} + \tilde{\delta})(1-\tilde{\delta}+\tilde{\gamma}) + \tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})} \approx (1-\tilde{\kappa})(\tilde{r}_t - \tilde{\gamma})$$

so that we may write (B.25c) as

$$\tilde{r}_{t+1}^f = \tilde{r}_t^f + \tilde{\kappa}(\tilde{\gamma} - \tilde{\delta} - \tilde{r}_t^f) - \tfrac{1}{2}\tilde{\kappa}\frac{(\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2} + \frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}\tilde{\eta}\epsilon_{A,t+1}$$

B.6 Summary of the two empirical specifications

In what follows, we use the empirical specifications where we may use financial market data for identification and estimation of structural parameters. For the continuous-time version we use the closed-form solution $C_t = \rho K_t$ together with the nonlinear equilibrium dynamics. For the discrete-time version we have approximately $\ln(C_t/K_t) = \ln \hat{C} + (\tilde{r}_t - \tilde{\gamma})/(1-\tilde{\delta}+\tilde{\gamma})$ together with the log-linear equilibrium dynamics.

First, based on quarterly data, the consumption (Euler) equations are

$$\ln(C_t/C_{t-1}) = \ln\tilde{\beta} + \tilde{r}^f_{t-1} + \tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2+\tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2} + \frac{\tilde{\sigma}}{\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})}\epsilon_{K,t} + \frac{\tilde{\eta}}{1-\tilde{\delta}+\tilde{\gamma}}\epsilon_{A,t}$$

vs.

$$\ln(C_t/C_{t-\Delta}) = \int_{t-\Delta}^{t} r^f_v dv - \left(\rho - \tfrac{1}{2}\sigma^2\right)\Delta + \sigma(Z_t - Z_{t-\Delta})$$

Second, based on quarterly data, the output equations (resource constraints) yield

$$\begin{aligned}\ln(Y_t/Y_{t-1}) &= \ln\tilde{\beta} + \tilde{r}^f_{t-1} + \tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2+\tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2} - \frac{(1-\tilde{\delta})\tilde{\kappa}}{(1-\tilde{\kappa})\tilde{\gamma}}\left(\tilde{r}^f_{t-1} - \tilde{\gamma} + \tilde{\delta} + \tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2+\tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2}\right)\\ &\quad + \frac{\tilde{\sigma}}{\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})}\epsilon_{K,t} + \frac{\tilde{\eta}}{\tilde{\gamma}}\epsilon_{A,t}\end{aligned}$$

vs.

$$\begin{aligned}\ln(Y_t/Y_{t-\Delta}) &= \int_{t-\Delta}^{t} r^f_v dv + \kappa\gamma\int_{t-\Delta}^{t} 1/(r^f_v+\delta+\sigma^2)dv - \tfrac{1}{2}\eta^2\int_{t-\Delta}^{t} 1/(r^f_v+\delta+\sigma^2)^2 dv\\ &\quad - \left(\kappa+\rho-\tfrac{1}{2}\sigma^2\right)\Delta + \int_{t-\Delta}^{t}\eta/(r^f_v+\delta+\sigma^2)dB_v + \sigma(Z_t - Z_{t-\Delta})\end{aligned}$$

Third, based on quarterly data, the interest rate equations (Vasicek specifications) read

$$\tilde{r}^f_t = (1-\tilde{\kappa})\tilde{r}^f_{t-1} + \tilde{\kappa}\left(\tilde{\gamma} - \tilde{\delta} - \tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2+\tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2}\right) + \frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}\tilde{\eta}\epsilon_{A,t}$$

vs.

$$r^f_t = e^{-\kappa\Delta}r^f_{t-\Delta} + (1-e^{-\kappa\Delta})(\gamma-\delta-\sigma^2) + \eta e^{-\kappa\Delta}\int_{t-\Delta}^{t} e^{\kappa(v-(t-\Delta))}dB_v$$

Note that r_t denotes the annual interest rate whereas $\tilde{r}_t$ denotes the periodic interest rate.

To summarize, the empirical specification comprises

$$
\begin{aligned}
\ln(C_t/C_{t-1}) &= \ln\tilde{\beta} + \tilde{r}^f_{t-1} + \tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2} + \frac{\tilde{\sigma}}{\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})}\epsilon_{K,t} + \frac{\tilde{\eta}}{1-\tilde{\delta}+\tilde{\gamma}}\epsilon_{A,t} \qquad \text{(B.26a)}\\
\ln(Y_t/Y_{t-1}) &= \ln\tilde{\beta} + \tilde{r}^f_{t-1} + \tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2} - \frac{(1-\tilde{\delta})\tilde{\kappa}}{(1-\tilde{\kappa})\tilde{\gamma}}\left(\tilde{r}^f_{t-1} - \tilde{\gamma} + \tilde{\delta} + \tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2}\right)\\
&\quad + \frac{\tilde{\sigma}}{\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})}\epsilon_{K,t} + \frac{\tilde{\eta}}{\tilde{\gamma}}\epsilon_{A,t} \qquad \text{(B.26b)}\\
\tilde{r}^f_t &= (1-\tilde{\kappa})\tilde{r}^f_{t-1} + \tilde{\kappa}\left(\tilde{\gamma} - \tilde{\delta} - \tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2}\right) + \frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}\tilde{\eta}\epsilon_{A,t} \qquad \text{(B.26c)}
\end{aligned}
$$

or

$$
\begin{aligned}
\ln(C_t/C_{t-1}) &= \ln\tilde{\beta} + \tilde{r}^f_{t-1} + \mathbb{C}_0 + \varepsilon_{C,t}\\
\ln(Y_t/Y_{t-1}) &= \ln\tilde{\beta} + \tilde{r}^f_{t-1} + \mathbb{C}_0 - \mathbb{C}_2\left(\tilde{r}^f_{t-1} - \mathbb{C}_1\right) + \varepsilon_{Y,t}\\
\tilde{r}^f_t &= (1-\tilde{\kappa})\tilde{r}^f_{t-1} + \tilde{\kappa}\mathbb{C}_1 + \varepsilon_{r,t}
\end{aligned}
$$

where

$$
\mathbb{C}_2 \equiv \frac{(1-\tilde{\delta})\tilde{\kappa}}{(1-\tilde{\kappa})\tilde{\gamma}}, \quad \mathbb{C}_1 \equiv \tilde{\gamma} - \tilde{\delta} - \mathbb{C}_0, \quad \mathbb{C}_0 \equiv \tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2}
$$

and

$$
\begin{aligned}
\varepsilon_{C,t} &= \frac{\tilde{\sigma}}{\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})}\epsilon_{K,t} + \frac{\tilde{\eta}}{1-\tilde{\delta}+\tilde{\gamma}}\epsilon_{A,t}, \qquad \text{(B.27a)}\\
\varepsilon_{Y,t} &= \frac{\tilde{\sigma}}{\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})}\epsilon_{K,t} + \frac{\tilde{\eta}}{\tilde{\gamma}}\epsilon_{A,t} \qquad \text{(B.27b)}\\
\varepsilon_{r,t} &= \frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}\tilde{\eta}\epsilon_{A,t} \qquad \text{(B.27c)}
\end{aligned}
$$

We employ the following mapping of structural parameters $\phi = (\kappa, \gamma, \eta, \rho, \delta, \sigma)^\top$

$$
\begin{aligned}
\kappa &\simeq \tilde{\kappa} = 1 - e^{-\Delta\kappa}\\
\gamma &\simeq \tilde{\gamma} = \Delta\gamma\\
\eta &\simeq \tilde{\eta} = \Delta\eta\sqrt{(1-e^{-2\kappa\Delta})/(2\kappa)}\\
\rho &\simeq \tilde{\beta} = e^{-\Delta\rho}\\
\delta &\simeq \tilde{\delta} = 1 - e^{-\Delta\delta}\\
\sigma &\simeq \tilde{\sigma} = \Delta^{1/2}\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})\sigma
\end{aligned}
$$

in which $\Delta = 1/12$ for monthly data, $\Delta = 1/4$ for quarterly data.

B.7 Residual Covariance Matrix and MEF

From system (B.26) we obtain the fitted residual covariance matrix

$$\hat{\Sigma} \equiv \begin{pmatrix} \hat{\Sigma}_{CC} & \hat{\Sigma}_{CY} & \hat{\Sigma}_{Cr} \\ \hat{\Sigma}_{YC} & \hat{\Sigma}_{YY} & \hat{\Sigma}_{Yr} \\ \hat{\Sigma}_{rC} & \hat{\Sigma}_{rY} & \hat{\Sigma}_{rr} \end{pmatrix}$$

where

$$\hat{\Sigma}_{CC} = \frac{\tilde{\sigma}^2 + (\tilde{\beta}\tilde{\eta})^2}{(\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma}))^2}, \quad \hat{\Sigma}_{CY} = \frac{\tilde{\sigma}^2}{(\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma}))^2} + \frac{\tilde{\eta}^2}{\tilde{\gamma}(1-\tilde{\delta}+\tilde{\gamma})}, \quad \hat{\Sigma}_{Cr} = \frac{\tilde{\eta}^2(1-\tilde{\kappa})}{(1-\tilde{\delta}+\tilde{\gamma})^2}$$

$$\hat{\Sigma}_{YY} = \frac{\tilde{\sigma}^2}{(\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma}))^2} + \frac{\tilde{\eta}^2}{\tilde{\gamma}^2}, \quad \hat{\Sigma}_{Yr} = \frac{\tilde{\eta}^2(1-\tilde{\kappa})}{\tilde{\gamma}(1-\tilde{\delta}+\tilde{\gamma})}, \quad \hat{\Sigma}_{rr} = \frac{\tilde{\eta}^2(1-\tilde{\kappa})^2}{(1-\tilde{\delta}+\tilde{\gamma})^2}$$

which we may use for better identification of structural model parameters.

Let ϕ denote the parameter vector of interest and let $m_t = m_t(\phi)$ denote the 3-vector of martingale increments generated by the model, expressed in terms of data and parameters. Specifically, we let $m_t = \varepsilon_t = (\varepsilon_{C,t}, \varepsilon_{Y,t}, \varepsilon_{r,t})$ in (B.26) be a martingale difference sequence,

$$m_t = \begin{pmatrix} \ln(C_t/C_{t-1}) - \tilde{r}^f_{t-1} - (\ln\tilde{\beta} + \mathbb{C}_0) \\ \ln(Y_t/Y_{t-1}) - \tilde{r}^f_{t-1} - (\ln\tilde{\beta} + \mathbb{C}_0) + \mathbb{C}_2(\tilde{r}^f_{t-1} - \mathbb{C}_1) \\ \tilde{r}^f_t - (1-\tilde{\kappa})\tilde{r}^f_{t-1} - \tilde{\kappa}\mathbb{C}_1 \end{pmatrix} \tag{B.28}$$

The optimal weights are given by

$$w_t = \psi_t^\top (\Psi_t)^{-1}$$

where Ψ_t is the conditional variance of the vector martingale increment,

$$\Psi_t = Var_{t-1}(m_t) = E_{t-1}(m_t m_t^\top)$$

and ψ_t the conditional mean of its parameter derivative

$$\psi_t = E_{t-1}(\partial m_t / \partial \phi^\top).$$

Here, the conditional variance of the vector martingale increment is constant and given by the residual covariance matrix, $\Psi_t = \hat{\Sigma}$. Using the martingale increments (B.28), we get the derivatives $(\partial m_t/\partial\phi^\top)^\top$ with respect to the parameter vector $\phi = (\tilde{\kappa}, \tilde{\gamma}, \tilde{\eta}, \tilde{\beta}, \tilde{\delta}, \tilde{\sigma})^\top$,

$$\begin{pmatrix} 0 & (\partial\mathbb{C}_2/\partial\tilde{\kappa})(r^f_{t-1} - \mathbb{C}_1) & \tilde{r}^f_{t-1} - \mathbb{C}_1 \\ -(\partial\mathbb{C}_0/\partial\tilde{\gamma}) & -(\partial\mathbb{C}_0/\partial\tilde{\gamma}) + (\partial\mathbb{C}_2/\partial\tilde{\gamma})\tilde{r}^f_{t-1} - (\partial(\mathbb{C}_1\mathbb{C}_2)/\partial\tilde{\gamma}) & -\tilde{\kappa}(1 - (\partial\mathbb{C}_0/\partial\tilde{\gamma})) \\ -(\partial\mathbb{C}_0/\partial\tilde{\eta}) & -(\partial\mathbb{C}_0/\partial\tilde{\eta}) + \mathbb{C}_2(\partial\mathbb{C}_0/\partial\tilde{\eta}) & \tilde{\kappa}(\partial\mathbb{C}_0/\partial\tilde{\eta}) \\ -1/\tilde{\beta} - (\partial\mathbb{C}_0/\partial\tilde{\beta}) & -1/\tilde{\beta} - (\partial\mathbb{C}_0/\partial\tilde{\beta}) + \mathbb{C}_2(\partial\mathbb{C}_0/\partial\tilde{\beta}) & \tilde{\kappa}(\partial\mathbb{C}_0/\partial\tilde{\beta}) \\ -(\partial\mathbb{C}_0/\partial\tilde{\delta}) & -(\partial\mathbb{C}_0/\partial\tilde{\delta}) + (\partial\mathbb{C}_2/\partial\tilde{\delta})\tilde{r}^f_{t-1} - (\partial(\mathbb{C}_1\mathbb{C}_2)/\partial\tilde{\delta}) & \tilde{\kappa}(1 + (\partial\mathbb{C}_0/\partial\tilde{\delta})) \\ -(\partial\mathbb{C}_0/\partial\tilde{\sigma}) & -(\partial\mathbb{C}_0/\partial\tilde{\sigma}) + \mathbb{C}_2(\partial\mathbb{C}_0/\partial\tilde{\sigma}) & \tilde{\kappa}(\partial\mathbb{C}_0/\partial\tilde{\sigma}) \end{pmatrix}$$

where

$$\partial\mathbb{C}_0/\partial\tilde{\gamma} = -\frac{(\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^3} = -2\mathbb{C}_0/(1-\tilde{\delta}+\tilde{\gamma}),$$

$$\partial\mathbb{C}_0/\partial\tilde{\eta} = \tilde{\eta}/(1-\tilde{\delta}+\tilde{\gamma})^2,$$

$$\partial\mathbb{C}_0/\partial\tilde{\beta} = -(\tilde{\sigma}/\tilde{\beta})^2/(\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma})^2),$$

$$\partial\mathbb{C}_0/\partial\tilde{\delta} = \frac{(\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^3} = 2\mathbb{C}_0/(1-\tilde{\delta}+\tilde{\gamma}),$$

$$\partial\mathbb{C}_0/\partial\tilde{\sigma} = \tilde{\sigma}/(\tilde{\beta}(1-\tilde{\delta}+\tilde{\gamma}))^2,$$

and

$$\partial\mathbb{C}_1/\partial\tilde{\gamma} = 1 - (\partial\mathbb{C}_0/\partial\tilde{\gamma})$$

$$\partial\mathbb{C}_1/\partial\tilde{\delta} = -1 - (\partial\mathbb{C}_0/\partial\tilde{\delta})$$

and

$$\partial\mathbb{C}_2/\partial\tilde{\kappa} = \frac{1-\tilde{\delta}}{(1-\tilde{\kappa})^2\tilde{\gamma}} = \mathbb{C}_2/(\tilde{\kappa}(1-\tilde{\kappa}))$$

$$\partial\mathbb{C}_2/\partial\tilde{\gamma} = -(1-\tilde{\delta})\tilde{\kappa}/((1-\tilde{\kappa})\tilde{\gamma}^2) = -\mathbb{C}_2/\tilde{\gamma}$$

$$\partial\mathbb{C}_2/\partial\tilde{\delta} = -\tilde{\kappa}/((1-\tilde{\kappa})\tilde{\gamma})$$

and

$$\partial(\mathbb{C}_1\mathbb{C}_2)/\partial\tilde{\gamma} = (\partial\mathbb{C}_1/\partial\tilde{\gamma})\mathbb{C}_2 + (\partial\mathbb{C}_2/\partial\tilde{\gamma})\mathbb{C}_1 = \mathbb{C}_2 - (\partial\mathbb{C}_0/\partial\tilde{\gamma})\mathbb{C}_2 - \mathbb{C}_1\mathbb{C}_2/\tilde{\gamma}$$

$$\partial(\mathbb{C}_1\mathbb{C}_2)/\partial\tilde{\delta} = (\partial\mathbb{C}_1/\partial\tilde{\delta})\mathbb{C}_2 + (\partial\mathbb{C}_2/\partial\tilde{\delta})\mathbb{C}_1$$

and with respect to the parameter vector $\phi = (\kappa, \gamma, \eta, \rho, \delta, \sigma)^\top$,

$$\begin{pmatrix} -(\partial\mathbb{C}_0/\partial\tilde{\eta})(\partial\tilde{\eta}/\partial\kappa) & \phi_{12} & \Delta e^{-\Delta\kappa}(\tilde{r}^f_{t-1} - \mathbb{C}_1) + \tilde{\kappa}(\partial\mathbb{C}_0/\partial\tilde{\eta})(\partial\tilde{\eta}/\partial\kappa) \\ -(\partial\mathbb{C}_0/\partial\tilde{\gamma})\Delta - (\partial\mathbb{C}_0/\partial\tilde{\sigma})\Delta^{3/2}\tilde{\beta}\sigma & \phi_{22} & -\Delta\tilde{\kappa}(1 - (\partial\mathbb{C}_0/\partial\tilde{\gamma}) - (\partial\mathbb{C}_0/\partial\tilde{\sigma})\Delta^{1/2}\tilde{\beta}\sigma) \\ -(\partial\mathbb{C}_0/\partial\tilde{\eta})(\tilde{\eta}/\eta) & \phi_{32} & \tilde{\kappa}(\partial\mathbb{C}_0/\partial\tilde{\eta})(\tilde{\eta}/\eta) \\ \Delta & \Delta & 0 \\ -((\partial\mathbb{C}_0/\partial\tilde{\delta}) - (\partial\mathbb{C}_0/\partial\tilde{\sigma})\Delta^{1/2}\tilde{\beta})\Delta e^{-\Delta\delta} & \phi_{52} & \tilde{\kappa}\Delta e^{-\Delta\delta}(1 + (\partial\mathbb{C}_0/\partial\tilde{\delta})) - \tilde{\kappa}(\partial\mathbb{C}_0/\partial\tilde{\sigma})\Delta^{1/2}\tilde{\beta}\sigma \\ -(\partial\mathbb{C}_0/\partial\tilde{\sigma})(\tilde{\sigma}/\sigma) & \phi_{62} & \tilde{\kappa}(\partial\mathbb{C}_0/\partial\tilde{\sigma})(\tilde{\sigma}/\sigma) \end{pmatrix}$$

where

$$\phi_{12} \equiv -(\partial\mathbb{C}_0/\partial\tilde{\eta})(\partial\tilde{\eta}/\partial\kappa) + (\partial\mathbb{C}_2/\partial\tilde{\kappa})\Delta e^{-\Delta\kappa} r^f_{t-1} + (\mathbb{C}_2(\partial\mathbb{C}_0/\partial\tilde{\eta})(\partial\tilde{\eta}/\partial\kappa) - \mathbb{C}_1(\partial\mathbb{C}_2/\partial\tilde{\kappa}))\Delta e^{-\Delta\kappa},$$

$$\phi_{22} \equiv -(\partial\mathbb{C}_0/\partial\tilde{\gamma}\Delta + \partial\mathbb{C}_0/\partial\tilde{\sigma}\Delta^{1/2}\tilde{\beta}\sigma) + (\partial\mathbb{C}_2/\partial\tilde{\gamma})\Delta\tilde{r}^f_{t-1} - (\partial(\mathbb{C}_1\mathbb{C}_2)/\partial\gamma),$$

$$\phi_{32} \equiv -(\partial\mathbb{C}_0/\partial\tilde{\eta})(\tilde{\eta}/\eta) + \mathbb{C}_2(\partial\mathbb{C}_0/\partial\tilde{\eta})(\tilde{\eta}/\eta),$$

$$\phi_{52} \equiv -(\partial\mathbb{C}_0/\partial\tilde{\delta})\Delta e^{-\Delta\delta} + (\partial\mathbb{C}_2/\partial\tilde{\delta})\Delta e^{-\Delta\delta}\tilde{r}^f_{t-1} - (\partial(\mathbb{C}_1\mathbb{C}_2)/\partial\delta),$$

$$\phi_{62} \equiv -(\partial \mathbb{C}_0/\partial \tilde{\sigma})(\tilde{\sigma}/\sigma) + \mathbb{C}_2(\partial \mathbb{C}_0/\partial \tilde{\sigma})(\tilde{\sigma}/\sigma),$$

and

$$\partial \tilde{\eta}/\partial \kappa = \tfrac{1}{2}\tilde{\eta}\left(2\Delta e^{-2\kappa\Delta}/(1-e^{-2\kappa\Delta}) - 1/\kappa\right),$$

and

$$\partial(\mathbb{C}_1\mathbb{C}_2)/\partial\gamma = (\partial\mathbb{C}_1/\partial\gamma)\mathbb{C}_2 + (\partial\mathbb{C}_2/\partial\gamma)\mathbb{C}_1 = \Delta\mathbb{C}_2 - (\partial\mathbb{C}_0/\partial\tilde{\gamma})\mathbb{C}_2 - (\partial\mathbb{C}_0/\partial\tilde{\sigma})\Delta^{3/2}\tilde{\beta}\sigma\mathbb{C}_2 - \mathbb{C}_1\mathbb{C}_2/\tilde{\gamma}\Delta,$$

$$\partial(\mathbb{C}_1\mathbb{C}_2)/\partial\delta = (-1 - (\partial\mathbb{C}_0/\partial\tilde{\delta}) + (\partial\mathbb{C}_0/\partial\tilde{\sigma})\Delta^{1/2}\tilde{\beta}\sigma)\Delta e^{-\Delta\delta}\mathbb{C}_2 + (\partial\mathbb{C}_2/\partial\tilde{\delta})\Delta e^{-\Delta\delta}\mathbb{C}_1.$$

B.8 Calibration of model parameters - periodic rates

Suppose that we want to parameterize the Ornstein-Uhlenbeck process and the first-order autoregressive process (with the discrete time process being at periodic rates)

$$dx_t = \kappa(\gamma - x_t)dt + \eta dB_t, \quad x_0 \text{ given} \quad \text{and} \quad \tilde{x}_{t+1} = \mathbb{C}_0 + \mathbb{C}_1\tilde{x}_t + \mathbb{C}_2\epsilon_{t+1}, \tag{B.29}$$

where $\tilde{x}_0 = \Delta x_0$, where $\Delta = 1/12$ for monthly and $\Delta = 1/4$ for quarterly observations, B_t a standard Brownian motion, $0 < \mathbb{C}_1 < 1$ and $\epsilon_t \sim \mathcal{N}(0,1)$.[5] The solutions are

$$x_t = e^{-\kappa t}x_0 + (1-e^{-\kappa t})\gamma + e^{-\kappa t}\eta\int_0^t e^{\kappa v}dB_v \quad \text{and} \quad \tilde{x}_t = \mathbb{C}_1^t\tilde{x}_0 + \mathbb{C}_1^t\sum_{i=1}^{t}\mathbb{C}_1^{-i}(\mathbb{C}_0 + \mathbb{C}_2\epsilon_i)$$

Let us calibrate $\mathbb{C}_i$, $i = 0,1,2$, given a parametric value for κ, γ and η at quarterly frequency, such that the expected value $\mathbb{E}_0(\Delta x_\Delta) = \mathbb{E}_0(\tilde{x}_1)$, the variance $Var_0(\Delta x_\Delta) = Var_0(\tilde{x}_1)$, and the mean of the asymptotic distribution $\mathbb{E}(\Delta x) = \mathbb{E}(\tilde{x})$ coincide. It it straightforward to show that $\mathbb{E}_0(\Delta x_\Delta) = \Delta e^{-\Delta\kappa}x_0 + \Delta(1-e^{-\Delta\kappa})\gamma$ and $\mathbb{E}_0(\tilde{x}_1) = \mathbb{C}_1\tilde{x}_0 + \mathbb{C}_0$. Moreover, $\mathbb{E}(\Delta x) = \Delta\gamma$ and $\mathbb{E}(\tilde{x}) = \mathbb{C}_0/(1-\mathbb{C}_1)$. This gives $\mathbb{C}_0 = \Delta\gamma(1-\mathbb{C}_1)$ in which $\mathbb{C}_1$ is pinned down by

$$\begin{aligned} \Delta e^{-\Delta\kappa}x_0 + \Delta\gamma - \Delta\gamma e^{-\Delta\kappa} &= \mathbb{C}_1\Delta x_0 + \Delta\gamma - \Delta\gamma\mathbb{C}_1 \\ \Leftrightarrow e^{-\Delta\kappa}(x_0 - \gamma) &= \mathbb{C}_1(x_0-\gamma) \\ \Leftrightarrow \mathbb{C}_1 &= e^{-\Delta\kappa} \end{aligned}$$

From the Itô isometry we get

$$Var_0(\Delta x_\Delta) = \Delta^2 e^{-2\kappa\Delta}\eta^2\int_0^\Delta e^{2\kappa v}dv = \Delta^2\frac{\eta^2}{2\kappa}(1-e^{-2\kappa\Delta})$$

whereas

$$Var_0(\tilde{x}_1) = \mathbb{C}_2^2 Var_0(\epsilon_1) = \mathbb{C}_2^2$$

[5]Note that $(1+r_t)^\Delta = 1 + \tilde{r}_t$ or $\Delta\ln(1+r_t) = \ln(1+\tilde{r}_t)$ and thus $\Delta r_t \approx \tilde{r}_t$.

Equating terms implies:

$$\mathbb{C}_2 = \Delta\eta\sqrt{(1-e^{-2\kappa\Delta})/(2\kappa)}$$

As an example, our DGP with $\Delta = 1/4$, $\kappa = 0.2$, $\gamma = 0.1$ and $\eta = 0.01$ corresponds to the discrete-time process with parameterization $\mathbb{C}_0 \approx 0.001$, $\mathbb{C}_1 \approx 0.951$, and $\mathbb{C}_2 \approx 0.001$, or

$$dx_t = 0.2(0.1 - x_t)dt + 0.01dB_t \quad \simeq \quad \tilde{x}_{t+1} = 0.001 + 0.951\tilde{x}_t + 0.001\epsilon_{t+1}$$

Observe that $B_\Delta - B_0 \sim \mathcal{N}(0,\Delta)$ so we may use $\epsilon_t \equiv \Delta^{-1/2}(B_t - B_{t-\Delta}) \sim \mathcal{N}(0,1)$ in order to match the size of the realized shocks. Economically, $\tilde{x}_t$ now matches the Vasicek interest rate dynamics of quarterly interest rates observed at the quarterly frequency,

$$\tilde{x}_{t+1} - \tilde{x}_t = \tilde{\kappa}(\tilde{\gamma} - \tilde{x}_t) + \tilde{\eta}\epsilon_{t+1}, \quad \tilde{x}_0 = \Delta x_0$$

where

$$\tilde{\gamma} \equiv \Delta\gamma, \quad \tilde{\kappa} \equiv 1 - e^{-\Delta\kappa}, \quad \tilde{\eta} \equiv \Delta\eta\sqrt{(1-e^{-2\kappa\Delta})/(2\kappa)}.$$

B.9 Calibration of model parameters - stochastic depreciation

Next we want to relate the dynamics of the discrete-time resource constraint

$$\ln(K_{t+1}/K_t) = \ln(1+\nu) + \frac{1}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t - \tilde{\gamma}) + \frac{\tilde{\sigma}}{1+\nu}\epsilon_{K,t+1}$$

to the corresponding continuous-time formulation

$$\ln(K_{t+\Delta}/K_t) = \int_t^{t+\Delta} r_s ds - (\delta + \rho + \tfrac{1}{2}\sigma^2)\Delta + \sigma(Z_{t+\Delta} - Z_t),$$

where $\epsilon_{K,t+1} = \Delta^{-1/2}(Z_{t+\Delta} - Z_t) \sim \mathcal{N}(0,1)$ so in order to get the same conditional moments

$$\frac{\tilde{\sigma}}{1+\nu}\Delta^{-1/2} = \sigma,$$

$$-(\delta - \gamma + \rho + \tfrac{1}{2}\sigma^2)\Delta = \ln\tilde{\beta} + \ln(1-\tilde{\delta}+\tilde{\gamma}).$$

It can be simplified to $(\gamma - \delta - \frac{1}{2}\sigma^2)\Delta = \ln(1+\tilde{\gamma}-\tilde{\delta})$ which corresponds to our definitions.

B.10 Asset pricing

From (B.25a) we obtain $E_t(\exp(-\ln(C_{t+1}/C_t)))$ as

$$E_t\left(\exp\left(-\left[\ln(1+\nu) + \frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t - \tilde{\gamma}) + \frac{\tilde{\sigma}}{1+\nu}\epsilon_{K,t+1} + \frac{\tilde{\eta}}{1-\tilde{\delta}+\tilde{\gamma}}\epsilon_{A,t+1}\right]\right)\right)$$

$$= \exp\left(-\ln(1+\nu) - \frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t - \tilde{\gamma}) + \tfrac{1}{2}\frac{\tilde{\sigma}^2}{(1+\nu)^2} + \tfrac{1}{2}\frac{\tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2}\right)$$

$$= \tilde{\beta}^{-1}(1-\tilde{\delta}+\tilde{\gamma})^{-1}\exp\left(-\frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t - \tilde{\gamma}) + \frac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2 + \tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2}\right).$$

Observe that from (B.16a) the one-period risk-free rate of some bond, r_t^f, which is determined at the end of period t for the following period $t+1$ must satisfy

$$\begin{aligned} 1+\tilde{r}_t^f &= \left(\tilde{\beta}E_t[\exp\left(-\ln(C_{t+1}/C_t)\right)]\right)^{-1} \\ &= (1-\tilde{\delta}+\tilde{\gamma})\exp\left(\frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t-\tilde{\gamma})-\tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2+\tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2}\right) \end{aligned}$$

and so

$$\tilde{r}_t^f \approx \tilde{\gamma}-\tilde{\delta}+\frac{1-\tilde{\kappa}}{1-\tilde{\delta}+\tilde{\gamma}}(\tilde{r}_t-\tilde{\gamma})-\tfrac{1}{2}\frac{(\tilde{\sigma}/\tilde{\beta})^2+\tilde{\eta}^2}{(1-\tilde{\delta}+\tilde{\gamma})^2} \tag{B.30}$$

and the expected risk premium over net capital rewards is $\frac{1}{2}((\tilde{\sigma}/\tilde{\beta})^2+\tilde{\eta}^2)/(1-\tilde{\delta}+\tilde{\gamma})^2$.

C MEF extensions, and MEF with five conditional moment restrictions

C.1 AK-Vasicek-RS model

In the case of the AK-Vasicek model with regime switching (cf. Section 3.3.2), the system of equilibrium dynamics reads[6]

$$\begin{aligned} d\ln C_t &= \left(r_t-\rho-\delta-\tfrac{1}{2}\sigma^2\right)dt+\sigma dZ_t, && \text{(C.1a)}\\ d\ln Y_t &= \left(\kappa\gamma/r_t-\tfrac{1}{2}(\eta_t/r_t)^2+r_t-\kappa-\rho-\delta-\tfrac{1}{2}\sigma^2\right)dt+\eta_t/r_t dB_t+\sigma dZ_t, && \text{(C.1b)}\\ dr_t &= \kappa(\gamma-r_t)dt+\eta_t dB_t, && \text{(C.1c)}\\ d\eta_t &= (\eta_l-\eta_h)dq_{1,t}+(\eta_h-\eta_l)dq_{2,t}. && \text{(C.1d)} \end{aligned}$$

Using system (C.1) and the equilibrium asset-pricing condition (14), we obtain

$$\begin{aligned} \ln(C_t/C_{t-\Delta})-\int_{t-\Delta}^t r_v^f dv &= -\left(\rho-\tfrac{1}{2}\sigma^2\right)\Delta+\varepsilon_{C,t}, && \text{(C.2a)}\\ \ln(Y_t/Y_{t-\Delta})-\int_{t-\Delta}^t r_v^f dv &= \kappa\gamma\int_{t-\Delta}^t 1/(r_v^f+\delta+\sigma^2)dv-\tfrac{1}{2}\int_{t-\Delta}^t \eta_v^2/(r_v^f+\delta+\sigma^2)^2 dv \\ &\quad -\left(\kappa+\rho-\tfrac{1}{2}\sigma^2\right)\Delta+\varepsilon_{Y,t}, && \text{(C.2b)}\\ r_t^f &= e^{-\kappa\Delta}r_{t-\Delta}^f+(1-e^{-\kappa\Delta})(\gamma-\delta-\sigma^2)+\varepsilon_{r,t}, && \text{(C.2c)}\\ \eta_t &= \eta_{t-\Delta}+(\eta_l-\eta_h)\int_{t-\Delta}^t(\phi_1(\eta_v)-\phi_2(\eta_v))dv+\varepsilon_{\eta,t}. && \text{(C.2d)} \end{aligned}$$

[6]It can be shown that the analytical solution $C_t=\rho K_t$ is not affected by the presence of regime switches such that the relation between the risk-free rate and the rental rate of capital is still given by (14).

with martingale increments given by

$$\varepsilon_{C,t} = \sigma(Z_t - Z_{t-\Delta}), \tag{C.3a}$$

$$\varepsilon_{Y,t} = \int_{t-\Delta}^{t} \eta_v/(r_v^f + \delta + \sigma^2)dB_v + \sigma(Z_t - Z_{t-\Delta}), \tag{C.3b}$$

$$\varepsilon_{r,t} = e^{-\kappa\Delta} \int_{t-\Delta}^{t} \eta_v e^{\kappa(v-(t-\Delta))} dB_v, \tag{C.3c}$$

$$\varepsilon_{\eta,t} = (\eta_l - \eta_h) \int_{t-\Delta}^{t} (dq_{1,v} - \phi_1(\eta_v)dv - dq_{2,v} + \phi_2(\eta_v)dv). \tag{C.3d}$$

We let $m_t = \varepsilon_t = (\varepsilon_{C,t}, \varepsilon_{Y,t}, \varepsilon_{r,t})^\top$ from (C.3a)-(C.3c), so $\dim m = 3$. Clearly, m_t is a martingale difference sequence, and from system (C.2) we have that in terms of data and parameters

$$m_t = \begin{pmatrix} \ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^{t} r_v^f dv + \left(\rho - \frac{1}{2}\sigma^2\right)\Delta \\ \ln(Y_t/Y_{t-\Delta}) - \int_{t-\Delta}^{t} r_v^f dv + \left(\kappa + \rho - \frac{1}{2}\sigma^2\right)\Delta - \kappa\gamma \int_{t-\Delta}^{t} 1/(r_v^f + \delta + \sigma^2)dv \\ + \frac{1}{2}\int_{t-\Delta}^{t} \eta_v^2/(r_v^f + \delta + \sigma^2)^2 dv \\ r_t^f - (1 - e^{-\kappa\Delta})(\gamma - \delta - \sigma^2) - e^{-\kappa\Delta} r_{t-\Delta}^f \end{pmatrix}, \tag{C.4}$$

where the integrals are approximated by Riemann sums over days between $t - \Delta$ and t.

Similarly to the case of latent variables, our procedure is to derive some proxy moments for estimation, say, $m_t^* = E\left(m_t|\mathcal{F}_{t-\Delta}\right)$, given by (46). We may also obtain $\Psi_{t,11} = \sigma^2\Delta$, $\Psi_{t,22} = E_{t-\Delta}(\int_{t-\Delta}^{t} \eta_v^2/(r_v^f+\delta+\sigma^2)^2 dv)+\sigma^2\Delta$, $\Psi_{t,33} = e^{-2\kappa\Delta}E_{t-\Delta}(\int_{t-\Delta}^{t} \eta_v^2 e^{2\kappa(v-(t-\Delta))}dv)$, $\Psi_{t,12} = \sigma^2\Delta$, $\Psi_{t,13} = 0$, and $\Psi_{t,23} = E_{t-\Delta}((\int_{t-\Delta}^{t} \eta_v/(r_v^f+\delta+\sigma^2)dB_v)(e^{-\kappa_1\Delta}\int_{t-\Delta}^{t} \eta_v e^{\kappa_1(v-(t-\Delta))}dB_v))$.We use Euler approximations for $\Psi_{t,22}$, $\Psi_{t,23}$, and $\Psi_{t,33}$ such that

$$\Psi_t = \begin{pmatrix} \sigma^2\Delta & \sigma^2\Delta & 0 \\ \sigma^2\Delta & \eta_{t-\Delta}^2\Delta/(r_{t-\Delta}^f + \delta + \sigma^2)^2 + \sigma^2\Delta & \eta_{t-\Delta}^2 e^{-\kappa_1\Delta}\Delta/(r_{t-\Delta}^f + \delta + \sigma^2) \\ 0 & \eta_{t-\Delta}^2 e^{-\kappa\Delta}\Delta/(r_{t-\Delta}^f + \delta + \sigma^2) & e^{-2\kappa\Delta}\Delta\eta_{t-\Delta}^2 \end{pmatrix}, \tag{C.5}$$

where in the estimation we simply replace η_t by its proxy η_t^*. Note that this is time-varying, i.e., MEF is strictly more efficient than GMM, and consistency of the parameter estimates is not affected since the approximations only enter the weights.

Using moments m_t^* given by (46), we get the derivatives $(\partial m_t^*/\partial\phi^\top)^\top$ with respect to the parameter vector $\phi = (\kappa, \gamma, \eta_l, \eta_h, \phi_{lh}, \phi_{hl}, \rho, \delta, \sigma)^\top$

$$\begin{pmatrix} 0 & \Delta - \gamma \int_{t-\Delta}^{t} 1/(r_v^f + \delta + \sigma^2) dv & \begin{matrix} -\Delta e^{-\kappa\Delta}(\gamma - \delta - \sigma^2) \\ +\Delta e^{-\kappa\Delta} r_{t-\Delta}^f \end{matrix} \\ 0 & -\kappa \int_{t-\Delta}^{t} 1/(r_v^f + \delta + \sigma^2) dv & -(1 - e^{-\kappa\Delta}) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \Delta & \Delta & 0 \\ 0 & \begin{matrix} \kappa\gamma \int_{t-\Delta}^{t} 1/(r_v^f + \delta + \sigma^2)^2 dv \\ -E\left(\int_{t-\Delta}^{t} \eta_v^2/(r_v^f + \delta + \sigma^2)^3 dv | \mathcal{F}_{t-\Delta}\right) \end{matrix} & 1 - e^{-\kappa\Delta} \\ -\sigma\Delta & E\left(\begin{matrix} -\sigma\Delta \\ +2\sigma\kappa\gamma \int_{t-\Delta}^{t} 1/(r_v^f + \delta + \sigma^2)^2 dv \\ -2\sigma \int_{t-\Delta}^{t} \eta_v^2/(r_v^f + \delta + \sigma^2)^3 dv \end{matrix} \middle| \mathcal{F}_{t-\Delta} \right) & 2\sigma(1 - e^{-\kappa\Delta}) \end{pmatrix}. \tag{C.6}$$

We also use an Euler approximation for the unknown integrals, such that $\psi_t^\top$ reads

$$\begin{pmatrix} 0 & \Delta - \gamma\Delta/(r_{t-\Delta}^f + \delta + \sigma^2) & \begin{matrix} -\Delta e^{-\kappa\Delta}(\gamma - \delta - \sigma^2) \\ +\Delta e^{-\kappa\Delta} r_{t-\Delta}^f \end{matrix} \\ 0 & -\kappa\Delta/(r_{t-\Delta}^f + \delta + \sigma^2) & -(1 - e^{-\kappa\Delta}) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \Delta & \Delta & 0 \\ 0 & \begin{matrix} \kappa\gamma\Delta/(r_{t-\Delta}^f + \delta + \sigma^2)^2 \\ -(\eta_{t-\Delta}^*)^2\Delta/(r_v^f + \delta + \sigma^2)^3 \end{matrix} & 1 - e^{-\kappa\Delta} \\ -\sigma\Delta & \begin{matrix} -\sigma\Delta + 2\sigma\kappa\gamma\Delta/(r_{t-\Delta}^f + \delta + \sigma^2)^2 \\ -2\sigma\Delta(\eta_{t-\Delta}^*)^2/(r_{t-\Delta}^f + \delta + \sigma^2)^3 \end{matrix} & 2\sigma(1 - e^{-\kappa\Delta}) \end{pmatrix}. \tag{C.7}$$

This completes the construction of the estimating function $M_T = \sum_t \psi_t^\top (\Psi_t)^{-1} m_t^*$.

C.2 AK-Vasicek-SV model

In the case of the AK-Vasicek model with stochastic volatility (cf. Section 3.3.3), the system of equilibrium dynamics reads[7]

$$
\begin{aligned}
d\ln C_t &= \left(r_t - \rho - \delta - \tfrac{1}{2}\sigma^2\right) dt + \sigma dZ_t, && \text{(C.8a)}\\
d\ln Y_t &= \left(\kappa_1\gamma_1/r_t - \tfrac{1}{2}(\eta_t/r_t)^2 + r_t - \kappa_1 - \rho - \delta - \tfrac{1}{2}\sigma^2\right) dt + \eta_t/r_t dB_t + \sigma dZ_t, && \text{(C.8b)}\\
dr_t &= \kappa_1(\gamma_1 - r_t)dt + \eta_t dB_t, && \text{(C.8c)}\\
d\log(\eta_t^2) &= \kappa_2(\gamma_2 - \log(\eta_t^2))dt + \xi dW_t. && \text{(C.8d)}
\end{aligned}
$$

Using system (C.8) and the equilibrium asset-pricing condition (14), we obtain

$$
\begin{aligned}
\ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^{t} r_v^f dv &= -\left(\rho - \tfrac{1}{2}\sigma^2\right)\Delta + \varepsilon_{C,t}, && \text{(C.9a)}\\
\ln(Y_t/Y_{t-\Delta}) - \int_{t-\Delta}^{t} r_v^f dv &= \kappa_1\gamma_1 \int_{t-\Delta}^{t} 1/(r_v^f + \delta + \sigma^2)dv - \tfrac{1}{2}\int_{t-\Delta}^{t} \eta_v^2/(r_v^f + \delta + \sigma^2)^2 dv\\
&\quad - \left(\kappa_1 + \rho - \tfrac{1}{2}\sigma^2\right)\Delta + \varepsilon_{Y,t}, && \text{(C.9b)}\\
r_t^f &= e^{-\kappa_1\Delta} r_{t-\Delta}^f + (1 - e^{-\kappa_1\Delta})(\gamma_1 - \delta - \sigma^2) + \varepsilon_{r,t}, && \text{(C.9c)}\\
\log(\eta_t^2) &= e^{-\kappa_2\Delta}\log(\eta_{t-\Delta}^2) + (1 - e^{-\kappa_2\Delta})\gamma_2 + \varepsilon_{\eta,t}, && \text{(C.9d)}
\end{aligned}
$$

with martingale increments given by

$$
\begin{aligned}
\varepsilon_{C,t} &= \sigma(Z_t - Z_{t-\Delta}), && \text{(C.10a)}\\
\varepsilon_{Y,t} &= \int_{t-\Delta}^{t} \eta_v/(r_v^f + \delta + \sigma^2)dB_v + \sigma(Z_t - Z_{t-\Delta}), && \text{(C.10b)}\\
\varepsilon_{r,t} &= e^{-\kappa_1\Delta}\int_{t-\Delta}^{t} \eta_v e^{\kappa_1(v-(t-\Delta))}dB_v, && \text{(C.10c)}\\
\varepsilon_{\eta,t} &= e^{-\kappa_2\Delta}\int_{t-\Delta}^{t} \xi e^{\kappa_2(v-(t-\Delta))}dW_v. && \text{(C.10d)}
\end{aligned}
$$

We let $m_t = \varepsilon_t = (\varepsilon_{C,t}, \varepsilon_{Y,t}, \varepsilon_{r,t}, \varepsilon_{\eta,t})^\top$ from (C.10a)-(C.10d), i.e., using four moments instead of three. Clearly, m_t is a martingale difference sequence, and from system (C.9) we have that in terms of data and parameters

$$
m_t = \begin{pmatrix}
\ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^{t} r_v^f dv + \left(\rho - \tfrac{1}{2}\sigma^2\right)\Delta\\
\ln(Y_t/Y_{t-\Delta}) - \int_{t-\Delta}^{t} r_v^f dv + \left(\kappa_1 + \rho - \tfrac{1}{2}\sigma^2\right)\Delta - \kappa_1\gamma_1\int_{t-\Delta}^{t} 1/(r_v^f + \delta + \sigma^2)dv\\
+\tfrac{1}{2}\int_{t-\Delta}^{t} \eta_v^2/(r_v^f + \delta + \sigma^2)^2 dv\\
r_t^f - (1 - e^{-\kappa_1\Delta})(\gamma_1 - \delta - \sigma^2) - e^{-\kappa_1\Delta} r_{t-\Delta}^f\\
\log(\eta_t^2) - (1 - e^{-\kappa_2\Delta})\gamma_2 - e^{-\kappa_2\Delta}\log(\eta_{t-\Delta}^2)
\end{pmatrix}, \qquad \text{(C.11)}
$$

[7]It can be shown that the analytical solution $C_t = \rho K_t$ is not affected by the presence of stochastic volatility such that the relation between the risk-free rate and the rental rate of capital is still given by (14).

where the integrals are approximated by Riemann sums over days between $t-\Delta$ and t.

Similarly to the case of latent interest rates, our procedure is to derive moments for estimation, say, $m_t^* = E\left(m_t|\mathcal{F}_{t-\Delta}\right)$, given by (48). We may also obtain $\Psi_{t,11} = \sigma^2\Delta$, $\Psi_{t,22} = E_{t-\Delta}(\int_{t-\Delta}^t \eta_v^2/(r_v^f+\delta+\sigma^2)^2 dv) + \sigma^2\Delta$, $\Psi_{t,33} = e^{-2\kappa_1\Delta}E_{t-\Delta}(\int_{t-\Delta}^t \eta_v^2 e^{2\kappa_1(v-(t-\Delta))})dv$, $\Psi_{t,44} = \xi^2(1-e^{-2\kappa_2\Delta})/(2\kappa_2)$, $\Psi_{t,12} = \sigma^2\Delta$, $\Psi_{t,13} = 0$, $\Psi_{t,14} = 0$, $\Psi_{t,23} = E_{t-\Delta}((\int_{t-\Delta}^t \eta_v/(r_v^f+\delta+\sigma^2)dB_v)(e^{-\kappa_1\Delta}\int_{t-\Delta}^t \eta_v e^{\kappa_1(v-(t-\Delta))}dB_v))$, $\Psi_{24}=0$, and $\Psi_{34}=0$. We use Euler approximations for the unknown integrals $\Psi_{t,22}$, $\Psi_{t,23}$ and $\Psi_{t,33}$, such that

$$\Psi_t = \begin{pmatrix} \sigma^2\Delta & \sigma^2\Delta & 0 & 0 \\ \sigma^2\Delta & \eta_{t-\Delta}^2\Delta/(r_{t-\Delta}^f+\delta+\sigma^2)^2+\sigma^2\Delta & \eta_{t-\Delta}^2 e^{-\kappa_1\Delta}\Delta/(r_{t-\Delta}^f+\delta+\sigma^2) & 0 \\ 0 & \eta_{t-\Delta}^2 e^{-\kappa_1\Delta}\Delta/(r_{t-\Delta}^f+\delta+\sigma^2) & e^{-2\kappa_1\Delta}\Delta\eta_{t-\Delta}^2 & 0 \\ 0 & 0 & 0 & \Psi_{t,44} \end{pmatrix}, \tag{C.12}$$

where in the estimation we simply replace η_t by its proxy η_t^*. Again, this is time-varying, and MEF strictly more efficient than GMM.

Using moments m_t given by (C.11), we get the derivatives $(\partial m_t/\partial\phi^\top)^\top$ with respect to the parameter vector $\phi = (\kappa_1,\gamma_1,\kappa_2,\gamma_2,\xi,\rho,\delta,\sigma)^\top$. We also use an Euler approximation for the unknown integrals, such that $\psi_t^\top$ reads

$$\begin{pmatrix} 0 & \Delta-\gamma_1\Delta/(r_{t-\Delta}^f+\delta+\sigma^2) & \begin{matrix}-\Delta e^{-\kappa_1\Delta}(\gamma_1-\delta-\sigma^2)\\ +\Delta e^{-\kappa_1\Delta}r_{t-\Delta}^f\end{matrix} & 0 \\ 0 & -\kappa_1\Delta/(r_{t-\Delta}^f+\delta+\sigma^2)dv & -(1-e^{-\kappa_1\Delta}) & 0 \\ 0 & 0 & 0 & \begin{matrix}-\Delta e^{-\kappa_2\Delta}\gamma_2\\ +\Delta e^{-\kappa_2\Delta}2\log(\eta_{t-\Delta}^*)\end{matrix} \\ 0 & 0 & 0 & -(1-e^{-\kappa_2\Delta}) \\ 0 & 0 & 0 & 0 \\ \Delta & \Delta & 0 & 0 \\ 0 & \begin{matrix}\kappa_1\gamma_1\Delta/(r_{t-\Delta}^f+\delta+\sigma^2)^2\\ -(\eta_{t-\Delta}^*)^2\Delta/(r_{t-\Delta}^f+\delta+\sigma^2)^3\end{matrix} & 1-e^{-\kappa_1\Delta} & 0 \\ -\sigma\Delta & \begin{matrix}-\sigma\Delta+2\sigma\kappa_1\gamma_1\Delta/(r_{t-\Delta}^f+\delta+\sigma^2)^2\\ -2\sigma(\eta_{t-\Delta}^*)^2\Delta/(r_{t-\Delta}^f+\delta+\sigma^2)^3\end{matrix} & 2\sigma(1-e^{-\kappa_1\Delta}) & 0 \end{pmatrix}. \tag{C.13}$$

This completes the construction of the estimating function $M_T = \sum_t \psi_t^\top(\Psi_t)^{-1}m_t^*$.

C.3 MEF with five conditional moment restrictions

The 5-vector $m_t = (\varepsilon_{C,t}, \varepsilon_{Y,t}, \varepsilon_{r,t}, \varepsilon^2_{C,t} - \sigma^2\Delta, \varepsilon^2_{r,t} - \eta^2(1-e^{-2\kappa\Delta})/(2\kappa))$ based on the error terms from (16) is clearly a martingale difference, given in terms of data and parameters as

$$m_t^{(5)} = \begin{pmatrix} \ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^t r_v^f dv + \left(\rho - \frac{1}{2}\sigma^2\right)\Delta \\ \ln(Y_t/Y_{t-\Delta}) - \int_{t-\Delta}^t r_v^f dv + \left(\kappa + \rho - \frac{1}{2}\sigma^2\right)\Delta - \kappa\gamma \int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2) dv \\ \qquad + \frac{1}{2}\eta^2 \int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)^2 dv \\ r_t^f - (1 - e^{-\kappa\Delta})(\gamma - \delta - \sigma^2) - e^{-\kappa\Delta} r_{t-\Delta}^f \\ \left(\ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^t r_v^f dv + \left(\rho - \frac{1}{2}\sigma^2\right)\Delta\right)^2 - \sigma^2\Delta \\ \left(r_t^f - (1 - e^{-\kappa\Delta})(\gamma - \delta - \sigma^2) - e^{-\kappa\Delta} r_{t-\Delta}^f\right)^2 - \eta^2(1 - e^{-2\kappa\Delta})/(2\kappa) \end{pmatrix}$$

or, by using the definition of three moment increments $m_t^{(3)}$ from (21),

$$m_t^{(5)} = \begin{pmatrix} m_t^{(3)} \\ \left(\ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^t r_v^f dv + \left(\rho - \frac{1}{2}\sigma^2\right)\Delta\right)^2 - \sigma^2\Delta \\ \left(r_t^f - (1 - e^{-\kappa\Delta})(\gamma - \delta - \sigma^2) - e^{-\kappa\Delta} r_{t-\Delta}^f\right)^2 - \eta^2(1 - e^{-2\kappa\Delta})/(2\kappa) \end{pmatrix} \tag{C.14}$$

which is equivalent to considering

$$m_t^{(5)} = \begin{pmatrix} \sigma(Z_t - Z_{t-\Delta}) \\ \int_{t-\Delta}^t \eta/(r_v^f + \delta + \sigma^2) dB_v + \sigma(Z_t - Z_{t-\Delta}) \\ \eta e^{-\kappa\Delta} \int_{t-\Delta}^t e^{\kappa(v-(t-\Delta))} dB_v \\ \sigma^2 (Z_t - Z_{t-\Delta})^2 - \sigma^2\Delta \\ \eta^2 e^{-2\kappa\Delta} \left(\int_{t-\Delta}^t e^{\kappa(v-(t-\Delta))} dB_v\right)^2 - \eta^2(1 - e^{-2\kappa\Delta})/(2\kappa) \end{pmatrix}$$

or

$$m_t^{(5)} = \begin{pmatrix} m_t^{(3)} \\ \sigma^2 (Z_t - Z_{t-\Delta})^2 - \sigma^2\Delta \\ \eta^2 e^{-2\kappa\Delta} \left(\int_{t-\Delta}^t e^{\kappa(v-(t-\Delta))} dB_v\right)^2 - \eta^2(1 - e^{-2\kappa\Delta})/(2\kappa) \end{pmatrix}. \tag{C.15}$$

To construct the MEF (27), we need the weights w_t in (29), which depend on the conditional mean of the parameter derivatives, ψ_t, and the conditional variance, Ψ_t, of m_t. We have the conditional variances $\Psi^{(5)}_{t,11} = \sigma^2\Delta$, $\Psi^{(5)}_{t,22} = \eta^2 E_{t-\Delta}(\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2)^2 dv) + \sigma^2\Delta$, and $\Psi^{(5)}_{t,33} = \eta^2(1 - e^{-2\kappa\Delta})/(2\kappa)$, $\Psi^{(5)}_{t,44} = 2\sigma^4\Delta^2$, and $\Psi^{(5)}_{t,55} = \eta^4 e^{-4\kappa\Delta} E_{t-\Delta}\left((\int_{t-\Delta}^t e^{\kappa(v-(t-\Delta))} dB_v)^4\right) - \frac{1}{4}\eta^4(1 - e^{-2\kappa\Delta})^2/\kappa^2$. Similarly, the conditional covariances are $\Psi^{(5)}_{t,12} = \sigma^2\Delta$, $\Psi^{(5)}_{t,13} = 0$, $\Psi^{(5)}_{t,14} = 0$, $\Psi^{(5)}_{t,15} = 0$, $\Psi^{(5)}_{t,23} = \eta^2 e^{-\kappa\Delta} E_{t-\Delta}\left((\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2) dB_v)(\int_{t-\Delta}^t e^{\kappa(v-(t-\Delta))} dB_v)\right)$, $\Psi^{(5)}_{t,24} = 0$, $\Psi^{(5)}_{t,25} = \eta^3 e^{-2\kappa\Delta} E_{t-\Delta}\left((\int_{t-\Delta}^t 1/(r_v^f + \delta + \sigma^2) dB_v)(\int_{t-\Delta}^t e^{\kappa(v-(t-\Delta))} dB_v)^2\right)$, $\Psi^{(5)}_{t,35} = \eta^3 e^{-3\kappa\Delta} E_{t-\Delta}\left((\int_{t-\Delta}^t e^{\kappa(v-(t-\Delta))} dB_v)^3\right)$, $\Psi^{(5)}_{t,34} = \Psi_{t,45} = 0$. We use Euler approximations for

$\Psi^{(5)}_{t,22}$, $\Psi^{(5)}_{t,55}$, $\Psi^{(5)}_{t,23}$, $\Psi^{(5)}_{t,25}$ and $\Psi^{(5)}_{t,35}$,

$$\Psi^{(5)}_t = \begin{pmatrix} \sigma^2\Delta & \sigma^2\Delta & 0 & 0 & 0 \\ \sigma^2\Delta & \sigma^2\Delta + \eta^2\Delta/(r^f_{t-\Delta}+\delta+\sigma^2)^2 & \eta^2 e^{-\kappa\Delta}\Delta/(r^f_{t-\Delta}+\delta+\sigma^2) & 0 & 0 \\ 0 & \eta^2 e^{-\kappa\Delta}\Delta/(r^f_{t-\Delta}+\delta+\sigma^2) & \frac{1}{2}\eta^2(1-e^{-2\kappa\Delta})/\kappa & 0 & 0 \\ 0 & 0 & 0 & 2\sigma^4\Delta^2 & 0 \\ 0 & 0 & 0 & 0 & \Psi^{(5)\prime}_{t,55} \end{pmatrix}$$

where $\Psi^{(5)\prime}_{t,55} = 3\eta^4 e^{-4\kappa\Delta}\Delta^2 - \frac{1}{4}\eta^4(1-e^{-2\kappa\Delta})^2/\kappa^2$, or using $\Psi^{(3)}_t$ from (41),

$$\Psi^{(5)}_t = \begin{pmatrix} \Psi^{(3)}_t & & 0_{3\times 2} \\ 0_{2\times 3} & 2\sigma^4\Delta^2 & 0 \\ & 0 & 3\eta^4 e^{-4\kappa\Delta}\Delta^2 - \frac{1}{4}\eta^4(1-e^{-2\kappa\Delta})^2/\kappa^2 \end{pmatrix}. \tag{C.16}$$

Again, $\Psi^{(5)}_t$ is time-varying, i.e., this is a conditionally heteroskedastic case, and optimal MEF is strictly more efficient than GMM. Consistency and the expression for the asymptotic variance are unaffected by our approximations because they enter only in the weights (29). Using martingale increments (C.14), we get the derivatives $(\partial m^{(5)}_t(\phi)/\partial\phi^\top)^\top$ with respect to the parameter vector $\phi = (\kappa, \gamma, \eta, \rho, \delta, \sigma)^\top$, such that $(\psi^{(5)}_t)^\top$ reads

$$\begin{pmatrix} & 0 & \frac{1}{2}\eta^2(1-e^{-2\kappa\Delta})/\kappa^2 - \eta^2\Delta e^{-2\kappa\Delta}/\kappa \\ & 0 & 0 \\ (\psi^{(3)}_t)^\top & 0 & -\eta(1-e^{-2\kappa\Delta})/\kappa \\ & 0 & 0 \\ & 0 & 0 \\ & -2\sigma\Delta & 0 \end{pmatrix}, \tag{C.17}$$

with $(\psi^{(3)}_t)^\top$ from (44), and where we use the fact that

$$\psi^{(5)}_{t,44} = 2E_{t-\Delta}\left(\ln(C_t/C_{t-\Delta}) - \int_{t-\Delta}^{t} r^f_v dv + (\rho - \tfrac{1}{2}\sigma^2)\Delta\right)\Delta = 0.$$

This completes the construction of the martingale estimating function for five conditional moment restrictions $M^{(5)}_T = \sum_t (\psi^{(5)}_t)^\top (\Psi^{(5)}_t)^{-1} m^{(5)}_t$.

D Additional Simulation Evidence

D.1 Simulation results: MEF extensions

In this section we present simulation results for two possible MEF extensions, namely, accommodating missing data points in the mixed-frequency approach (MF-MEF), and latent variables using the simulation-based approach (SMEF). Table D1 provides the results for the simulation study of the AK-Vasicek model for the case of truly missing data. In the first

Table D1: Simulation Study – Latent Short Rate and Mixed Frequency

The table reports output of a simulation study of the accuracy of the structural model parameters estimated using the latent short rate and mixed-frequency MEF approaches for the AK-Vasicek model, SMEF (Latent Short Rate) and MF-MEF, respectively. For 1,000 replications, we generate 25 years of data from the underlying data generating process (DGP) and apply our estimation strategy. We show the median estimate, and provide the interquartile range below it. For completeness we include the MEF estimates from Table 1.

Parameter Estimates from Simulation Study – SMEF (Latent Short Rate) and MF-MEF						
		Monthly Data		Quarterly Data		Mixed Frequency
	DGP	MEF	SMEF	MEF	SMEF	MF-MEF
κ	0.2	0.354	0.355	0.353	0.363	0.360
		0.284	0.280	0.305	0.290	0.290
γ	0.1	0.099	0.099	0.099	0.107	0.099
		0.013	0.012	0.013	0.020	0.013
η	0.01	0.010	0.010	0.010	0.010	0.010
		0.001	0.001	0.001	0.002	0.001
ρ	0.03	0.030	0.030	0.030	0.032	0.030
		0.006	0.002	0.006	0.005	0.006
δ	0.05	0.050	0.051	0.050	0.055	0.050
		0.002	0.005	0.003	0.013	0.002
σ	0.02	0.023	0.021	0.025	0.021	0.022
		0.005	0.003	0.010	0.006	0.005

column we list the parameter values as they are used in the data generating process (DGP), in column 3 the SMEF estimates obtained on simulated monthly data, in column 5 the SMEF estimates for the simulated quarterly data, and in column 6 the MF-MEF estimates for the mixed-frequency data. The interest rate data are missing in the SMEF cases, and two of every three monthly output observations are missing in the MF-MEF case. For comparison, we also replicate the complete data MEF results from Table 1 in columns 2 and 4, respectively, using monthly and quarterly data. For the case of observed data (consumption, output and interest rate), but with latent volatility, Table D2 provides the results for regime switching (Panel A) and stochastic volatility (Panel B). As before, we provide the median estimate of each parameter, and below the interquartile range of the 1,000 estimates.

For the latent variable extension, case (i) from Section 3.3, we compare SMEF (columns 3 and 5 of Table 2) with MEF estimates (columns 2 and 4 of Table D1). We find that the latent variable case is as good as the observed short rate process. At both the monthly and the quarterly observation frequency, the point estimates and interquartile ranges are estimated remarkably close to the DGP values and are comparable with the MEF figures, with slightly smaller interquartile ranges in the SMEF approach for ρ and σ. This suggests that the model-consistent interest rate proxy $r_t^* = \rho Y_t / C_t$ is particularly fortunate in the AK-Vasicek model. Of course, these findings hold true only if the data were simulated from the correct model. This fact allows us to run model-specification checks on the empirical data at hand. The simulated short rate process can actually be compared with some observed proxies (see also the discussion in Section 5.3).

Table D2: Simulation Study – Regime Switching and Stochastic Volatility

The table reports output of a simulation study of the accuracy of the structural model parameters estimated using the MEF approaches for the AK-Vasicek model with regime switching for the interest rate volatility (Panel A) and with latent stochastic volatility (Panel B). We use three strategies in the latter case for identifying ξ. First, we estimate it along with the other parameters ("MEF"). Second we fix it at the known value ("Fix ξ"). Third, we estimate it separately by looking at the residuals from an autoregressive process for the proxied volatility series and plug this value into the MEF procedure ("Proxy ξ"). For 1,000 replications, we generate 25 years of data from the underlying data generating process (DGP) and apply our estimation strategy. We show the median estimate, and provide the interquartile range below it.

Panel A: Simulation Study – Regime Switching (η_l and η_h)
MEF (Latent Short Rate Volatility)

		Monthly	Quarterly		Monthly	Quarterly
	DGP	MEF	MEF	DGP	MEF	MEF
κ	0.2	0.045	0.032	0.5	0.542	0.587
		0.090	0.128		0.388	0.423
γ	0.1	0.081	0.094	1	1.002	1.009
		0.085	0.092		0.086	0.134
η_l	0.005	0.019	0.016	0.1	0.113	0.118
		0.011	0.013		0.037	0.097
η_h	0.02	0.022	0.023	0.25	0.240	0.225
		0.009	0.019		0.064	0.106
ϕ_{lh}	1.1	1.075	1.034	1	1.045	1.327
		0.533	1.058		0.967	9.023
ϕ_{hl}	1.5	1.403	1.176	5	4.304	3.567
		0.756	1.191		3.471	47.847
ρ	0.03	0.030	0.031	0.03	0.030	0.029
		0.006	0.007		0.006	0.008
δ	0.05	0.080	0.091	0.05	0.049	0.050
		0.104	0.176		0.002	0.033
σ	0.02	0.021	0.021	0.02	0.023	0.027
		0.009	0.029		0.013	0.062

Panel B: Simulation Study – Stochastic Volatility (Latent η_t)
MEF (Latent Short Rate Volatility)

		Monthly			Quarterly		
	DGP	MEF	Fix ξ	Proxy ξ	MEF	Fix ξ	Proxy ξ
κ_1	0.2	0.243	0.250	0.251	0.244	0.268	0.269
		0.201	0.215	0.208	0.253	0.285	0.289
γ_1	0.1	0.099	0.102	0.101	0.103	0.113	0.113
		0.079	0.080	0.080	0.095	0.136	0.134
κ_2	2	2.065	2.173	2.180	0.267	1.299	1.310
		1.084	0.706	0.712	1.588	1.631	1.647
γ_2	-10	−10.033	−10.034	−10.030	−9.891	−9.887	−9.892
		0.413	0.358	0.357	3.469	0.495	0.487
ξ	2.5	4.726	2.500	2.565	326.666	2.500	1.968
		156.150		0.165	6091.628		0.222
ρ	0.03	0.031	0.031	0.031	0.031	0.032	0.032
		0.006	0.006	0.006	0.007	0.007	0.007
δ	0.05	0.047	0.050	0.050	0.050	0.059	0.058
		0.073	0.077	0.075	0.089	0.119	0.110
σ	0.02	0.022	0.022	0.022	0.025	0.028	0.028
		0.027	0.032	0.035	0.052	0.060	0.059

For the latent variable extension, case (ii), we find that the identifiability of the structural parameters largely depends on the calibration of the DGP values (cf. Table D2, Panel A). This is intuitive because the embedded filter may have problems identifying the transition probabilities and/or the size of the two volatility regimes if the difference between them is negligible. One DGP in column 1 is taken roughly in line with the interest rate data, while another DGP in column 4 illustrates that performance improves if the difference between the two regimes is more pronounced. The regime-switching model also has strong implications for the estimate (and bias) of the mean-reversion parameter κ. The point estimates and interquartile ranges for parameters γ, ρ and σ are estimated remarkably close to DGP values. The upward bias in δ in columns 2 and 3 may be explained by a weak identification of η_l. For the case where the two regimes are well identified in columns 5 and 6, all parameter estimates are close to their DGP values. Overall, sampling data at a higher frequency works better.

For the latent variable extension, case (iii), we find that the parameter ξ, the variance of the stochastic volatility process, is weakly identified (cf. Table D2, Panel B). This explains the large interquartile range for the SMEF estimates (columns 2 and 5). Hence, in columns 3 and 7 we fix ξ at its DGP value. We then compare SMEF estimates (columns 4 and 7) with the benchmark estimates when ξ is known. We find that estimating ξ at the outset and using this value as a proxy for ξ works remarkably well. At both the monthly and the quarterly observation frequency, the point estimates and interquartile ranges are estimated remarkably close to DGP values and are comparable with the benchmark figures.

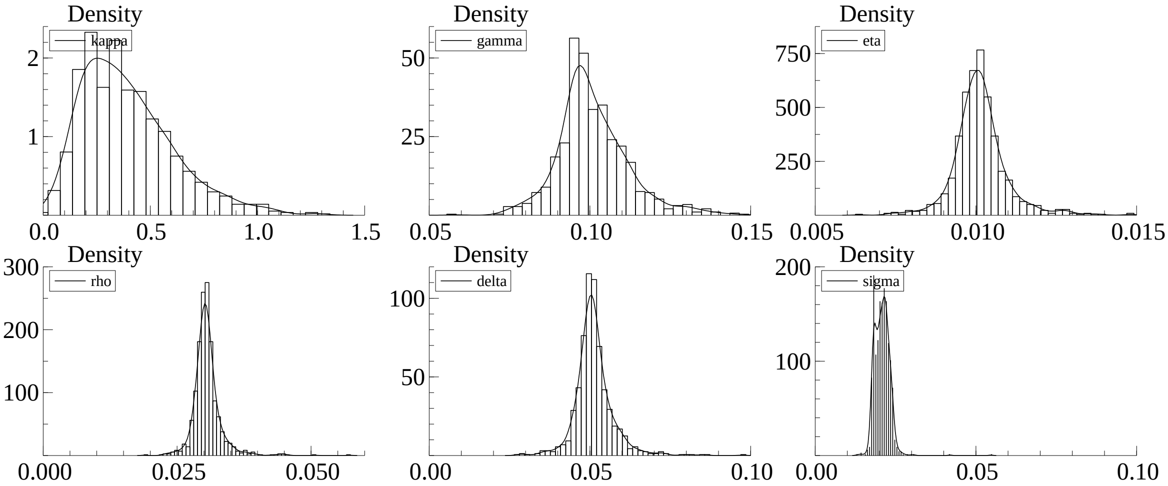
For the extension to mixed-frequency data, case (iv), we compare MF-MEF (column 6) with MEF estimates (columns 2 and 4). As one would expect, given the correct specification, for the case when output is replaced by model consistent predictions at intra-quarter periods the point estimates are remarkably close to the monthly estimates. Comparing the MF-MEF results to MEF, where consumption and output is observed at the quarterly frequency, we find that we gain better identification in σ, reflected by the smaller interquartile range.

In Figure D1 we provide the histograms of the 1,000 estimates that we obtain for the parameters using both the SMEF for monthly data and the MF-MEF approaches (Table D1). Comparing the histograms of SMEF to monthly MEF in Figure 1 (both Panel A) illustrates that ρ and σ are better identified in SMEF, which also is reflected by smaller interquartile ranges above, while the histogram is slightly more narrow for δ in the MEF approach. Similarly, comparing the histograms of MF-MEF (Panel B) to monthly and quarterly MEF, respectively, in Figure 1 (Panels A and B) shows that there is a small efficiency loss with respect to monthly data, but better identification of parameters is obtained relative to the results when estimates are obtained solely from quarterly data.

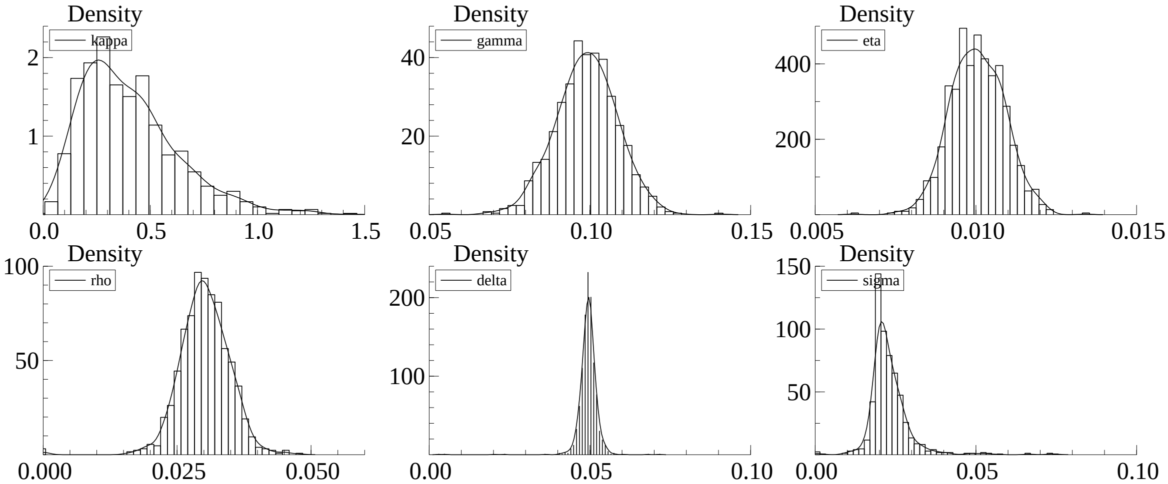
Figure D1: Simulation Study – Latent Short Rate and Mixed Frequency

The figure reports output of a simulation study of the accuracy of the structural model parameters estimated using simulated MEF and mixed-frequency approaches for the AK-Vasicek model, SMEF, and MF-MEF, respectively. For 1,000 replications, we generate 25 years of data from the underlying data generating process (DGP) and apply our estimation strategy. We plot the distribution of the estimates, in Panel A for the SMEF (Latent Short Rate) case based on monthly data and in Panel B for the MF-MEF approach.

(A) SMEF (Latent Short Rate) for Monthly Data



(B) MF-MEF



Overall, our MEF extensions work and are potentially more important for the case when we apply the methods to empirical data by the same reasons that motivated our extensions.

D.2 Robustness: Time invariance, high-frequency data, and the comparison to discrete time

In this section we present robustness simulation results which are particularly relevant for the estimation of continuous-time models. We want to provide answers to the following three questions: (i) Are the estimates time invariant? In theory, the continuous-time model is time invariant. However, different continuous-time processes may look identical if sampled at discrete points, which sometimes is referred to as the aliasing problem. This phenomenon may prevent unique identification of the parameters of the continuous-time stochastic process from equidistant discrete-time observations. Moreover, any temporal aggregation of the data may distort our parameter estimates. For these reasons, it seems important to examine to which extent our parameter estimates change with the observation frequency. (ii) Do the high-frequency data matter? So far, we only exploit the high-frequency property of the interest rate in the approximation of the integrals as Riemann sums. Hence, we want to examine to which extent the use of daily observations versus only considering the end-of-period figure helps to identify the parameters in our analysis. (iii) What happens if the true DGP is the continuous-time model and the researcher specifies a discrete-time model, then estimates that system to obtain parameter estimates. Is this problematic?

In order to examine (i), to which extent the parameter estimates change with the sampling frequency, we simulate monthly and quarterly data respectively, with the same number of observations for comparison. In Table D3 we compare the usual 25 years of monthly data to 75 years of simulated quarterly data (Panel A). As before, we provide the median estimate of each parameter, and below the interquartile range of the 1,000 estimates. The results show that the bias in the κ estimate is much smaller with quarterly data than if the data were sampled at monthly frequency. Moreover, the interquartile ranges are substantially smaller with quarterly data for all four estimation methods. This reveals that the time invariance property translates to all parameters of interest except the mean-reversion parameter κ. This upward bias, however, seems to diminish if quarterly data were used, provided the number of observations is sufficiently large (compare also to the results in Table 1).

Table D3: Robustness Simulations – Time invariance and High-frequency data

The table reports output of two simulation studies of the robustness of our estimation methods. In a first simulation (Panel A) we examine to which extent the parameter estimates change with the sampling frequency. To this end, we simulate the quarterly data set-up with the same number of observations as in the monthly set-up. Specifically, we compare the usual 25 years of monthly data to 75 years of quarterly data. In a second robustness simulation (Panel B), we examine what our results would look like without exploiting the availability of daily interest rates, using instead only used the end-of-period number. To this end, we simulate the data as usual, but only use the end-of-month and end-of-quarter short rate in our estimation, rather than the integrals. We show the accuracy of the structural model parameters estimated using OLS, FGLS-SUR-IV, GMM, and MEF for the AK-Vasicek model with three conditional moment restrictions. For 1,000 replications, we generate the data from the underlying data generating process (DGP) and apply our estimation strategy. We show the median estimate, and provide the interquartile range below it.

Panel A: Robustness Simulations – Time invariance

		Monthly Data				Quarterly Data (75 years)			
	DGP	OLS	FGLS-SUR-IV	GMM	MEF	OLS	FGLS-SUR-IV	GMM	MEF
κ	0.2	0.349	0.299	0.345	0.354	0.236	0.179	0.246	0.246
		0.286	0.134	0.345	0.284	0.116	0.058	0.145	0.120
γ	0.1	0.201	0.101	0.100	0.100	0.190	0.101	0.101	0.100
		0.036	0.013	0.014	0.013	0.023	0.008	0.009	0.008
η	0.01	0.083	0.008	0.010	0.010	0.065	0.007	0.010	0.010
		0.036	0.004	0.001	0.001	0.019	0.002	0.001	0.001
ρ	0.03	0.080	0.030	0.030	0.030	0.075	0.030	0.030	0.030
		0.015	0.006	0.007	0.006	0.009	0.003	0.004	0.003
δ	0.05	0.05	0.05	0.05	0.050	0.05	0.05	0.05	0.050
					0.002				0.002
σ	0.02	0.317	0.000	0.027	0.023	0.299	0.000	0.018	0.022
		0.040	<0.001	0.047	0.005	0.031	<0.001	0.052	0.006

Panel B: Robustness Simulations – Daily vs. Monthly and Quarterly Short Rate

		Monthly Data				Quarterly Data			
	DGP	OLS	FGLS-SUR-IV	GMM	MEF	OLS	FGLS-SUR-IV	GMM	MEF
κ	0.2	0.188	0.395	0.164	0.356	0.184	0.387	0.144	0.353
		0.444	0.260	0.164	0.286	0.444	0.287	0.137	0.305
γ	0.1	0.241	0.100	0.099	0.098	0.236	0.100	0.099	0.094
		0.205	0.013	0.017	0.012	0.195	0.013	0.018	0.013
η	0.01	0.077	0.009	0.010	0.010	0.075	0.009	0.010	0.010
		0.105	0.004	0.001	0.001	0.102	0.003	0.002	0.001
ρ	0.03	0.104	0.030	0.030	0.030	0.101	0.030	0.031	0.030
		0.095	0.006	0.006	0.006	0.091	0.006	0.007	0.006
δ	0.05	0.05	0.05	0.05	0.048	0.05	0.05	0.05	0.045
					0.002				0.004
σ	0.02	0.389	0.000	0.000	0.023	0.383	0.000	0.032	0.024
		0.435	0.011	0.038	0.005	0.427	<0.001	0.057	0.010

We also examine (ii), what our results would look like if we did not use the daily availability of interest rates, but only the end-of-period number. To this end we simulate the data as usual, but only use the end-of-month and end-of-quarter short rate in our estimation, rather than the integrals. This simply neglects all within-period dynamics. In Table D3 we show the results for monthly data and quarterly data (Panel B). Comparing to the results in Table 1 shows that neglecting within-period dynamics is not innocuous. The general pattern is that it comes at the cost of increasing inter-quartile ranges and changing parameter estimates. In particular, the estimate for the mean-reversion parameter κ changes substantially for OLS, FGLS-SUR-IV, and GMM. Moreover, we get into more severe identification problems for σ in the regression-based approaches and now also for GMM. In contrast, we observe only minor efficiency losses for the MEF approach. Here, we still provide some information about the dynamics of the stochastic process using the deterministic Taylor expansion (43). Table D4 shows that the use of high-frequency observations may be more important for different models and/or data. In particular, if the speed of mean reversion κ is high, the daily approximations of integrals of financial interest rate data are important for identification of the structural parameters of the macro model, such as the depreciation rate δ. For example, using DGP values $\kappa = 1$ and $\delta = 0.05$, the end-of-quarter approximation of integrals suggests that $\delta = 0$, and at the same time the high-frequency data approximation yields $\delta = 0.049$. These patterns suggest that both high-frequency data and/or more information about the within-period dynamics help identify the parameters of interest.

To examine (iii) we simulate the DGP from the continuous-time system (15) and then estimate the discrete time system (19). It turns out that while the model is misspecified, at a first glance, it seems to be a (surprisingly) good approximation and the structural parameters can be estimated from the simulated data (cf. Table D5). We may directly compare the results to Table 1. We find that the median MEF point estimates are similar in both approaches. A second look, however, reveals that the approximation has strong consequences on the identifiability of structural parameters. As shown by Canova and Sala (2009), many (linear) DSGE models share identification problems for (a subset of) model parameters. Figure D2, which shows the elasticity of the objective function to the parameter values for one draw of the simulated data, suggests that the continuous-time approach, where we use the nonlinear model, may help the identifiability of structural parameters.[8] The objective function is much steeper around parameter estimates in the continuous-time model (Panel A), which implies elasticities at several orders of magnitude higher in the continuous-time

[8] A nonlinear analysis of the discrete-time model requires solving (18) with some nonlinear approximation scheme. Along those lines, we would arrive at an (implicit) dynamic equilibrium system which no longer allows the application of our estimation methods. This is in contrast to our continuous-time approach where an explicit dynamic equilibrium system is obtained using the stochastic calculus.

Table D4: Robustness Simulations – High-frequency Data for Different DGPs

The table reports output of a simulation study of the accuracy of the structural model parameters estimated using the MEF approach for the AK-Vasicek model, to illustrate the benefits of high-frequency data. In each panel we show two types of results. First, estimation where the Riemann sum for the integral is replaced by the end of month ("EoMth") and end of quarter value ("EoQrt"). Second, the usual estimation using the Riemann sum approximation of the integral ("Daily"). For 1,000 replications, we generate 25 years of data from the underlying data generating process (DGP) and apply our estimation strategy. We show the median estimate, and provide the interquartile range below it. In Panel A (a) we replicate the results from Tables 1 and D3. In Panel A (b) we report results for a DGP setting with a relatively high value of η, whereas in Panel B (a) and (b) we report results for relatively high values of both κ and η.

Panel A: Robustness Simulations – Daily vs. Monthly and Quarterly Short Rate (high η)

	(a)	Monthly Data		Quarterly Data		(b)	Monthly Data		Quarterly Data	
	DGP	Daily	EoMth	Daily	EoQrt	DGP	Daily	EoMth	Daily	EoQrt
κ	0.2	0.354	0.356	0.353	0.353	0.2	0.228	0.212	0.228	0.189
		0.284	0.286	0.305	0.305		0.342	0.328	0.342	0.374
γ	0.1	0.099	0.098	0.099	0.094	0.5	0.521	0.521	0.521	0.589
		0.013	0.012	0.013	0.013		0.308	0.707	0.308	42.475
η	0.01	0.010	0.010	0.010	0.010	0.1	0.099	0.099	0.099	0.099
		0.001	0.001	0.001	0.001		0.009	0.009	0.009	0.021
ρ	0.03	0.030	0.030	0.030	0.030	0.03	0.031	0.031	0.031	0.032
		0.006	0.006	0.006	0.006		0.006	0.006	0.006	0.006
δ	0.05	0.050	0.048	0.050	0.045	0.05	0.043	0.037	0.043	0.028
		0.002	0.002	0.003	0.004		0.034	0.028	0.034	0.052
σ	0.02	0.023	0.023	0.025	0.024	0.02	0.022	0.022	0.022	0.020
		0.005	0.005	0.010	0.010		0.007	0.007	0.007	0.772

Panel B: Robustness Simulations – Daily vs. Monthly and Quarterly Short Rate (high κ and η)

	(a)	Monthly Data		Quarterly Data		(b)	Monthly Data		Quarterly Data	
	DGP	Daily	EoMth	Daily	EoQrt	DGP	Daily	EoMth	Daily	EoQrt
κ	0.5	0.754	0.740	0.754	0.724	1	1.129	1.123	1.129	1.162
		0.434	0.463	0.434	0.552		0.436	0.443	0.436	0.985
γ	0.5	0.538	0.532	0.538	0.541	0.5	0.498	0.476	0.498	0.452
		0.125	0.122	0.125	0.217		0.029	0.032	0.029	0.032
η	0.2	0.196	0.197	0.196	0.197	0.1	0.100	0.100	0.100	0.105
		0.019	0.019	0.019	0.042		0.007	0.006	0.007	0.018
ρ	0.03	0.031	0.031	0.031	0.032	0.03	0.031	0.031	0.031	0.031
		0.006	0.006	0.006	0.007		0.006	0.006	0.006	0.006
δ	0.05	0.048	0.038	0.048	0.011	0.05	0.049	0.026	0.049	0.000
		0.038	0.042	0.038	0.053		0.011	0.013	0.011	0.001
σ	0.02	0.022	0.022	0.022	0.020	0.02	0.022	0.022	0.022	0.001
		0.007	0.010	0.007	0.355		0.003	0.004	0.003	0.018

approach (Panel B). Moreover, within-period dynamics or the mixed-frequency property of macro and financial data can no longer be exploited (as discussed above). Hence, in models where the within-period dynamics are economically relevant and/or the nonlinearities are economically important, this will probably also show up in forecasting performance. Ultimately, it will be an empirical question whether these features matter in more elaborate models.

Our conclusion from the robustness analysis is that if parameters are well identified, a (log-linear) approximation of a more elaborate continuous-time model where no analytical solution is available seems a promising route and may be the best-practice approach. While it easily allows using mixed-frequency data, it keeps efficiency losses when estimating the model at a minimum. We leave further analysis of this for future research.

Table D5: Simulation Study – Monthly and Quarterly Data Discrete Time

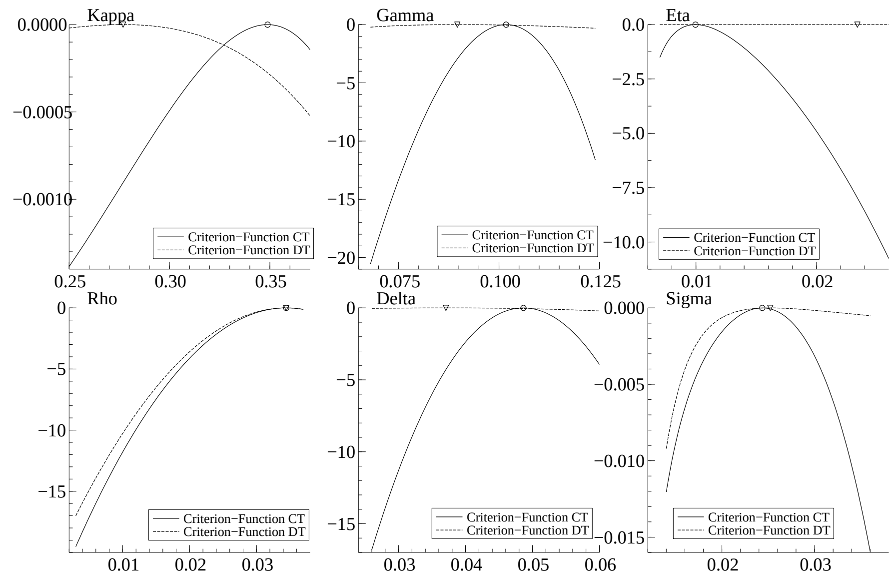
The table reports output of a simulation study of the accuracy of the structural model parameters estimated using the OLS, FGLS-SUR-IV, GMM, and MEF approaches to the AK-Vasicek model in Discrete Time. For 1,000 replications, we generate 25 years of data from the underlying continuous time data generating process (DGP) and apply our estimation strategy. We show the median estimate, and provide the interquartile range below it.

Parameter Estimates from Simulation Study – Monthly & Quarterly Data									
		Monthly Data				Quarterly Data			
	DGP	OLS	FGLS-SUR-IV	GMM	MEF	OLS	FGLS-SUR-IV	GMM	MEF
κ	0.200	0.361	0.135	0.365	0.282	0.366	0.114	0.385	0.335
		0.280	0.084	0.268	0.249	0.286	0.056	0.310	0.279
γ	0.100	0.103	0.101	0.099	0.096	0.107	0.103	0.108	0.105
		0.016	0.015	0.013	0.015	0.019	0.018	0.017	0.023
η	0.010	0.010	0.010	0.010	0.012	0.010	0.010	0.010	0.011
		0.010	0.010	0.000	0.014	0.011	0.010	0.000	0.008
ρ	0.030	0.034	0.033	0.031	0.030	0.039	0.034	0.039	0.036
		0.011	0.009	0.007	0.006	0.017	0.012	0.012	0.013
δ	0.050	0.050	0.050	0.050	0.047	0.050	0.050	0.050	0.050
					0.011				0.008
σ	0.020	0.072	0.034	0.020	0.020	0.134	0.067	0.133	0.109
		0.135	0.117	0.001	0.002	0.187	0.150	0.077	0.136

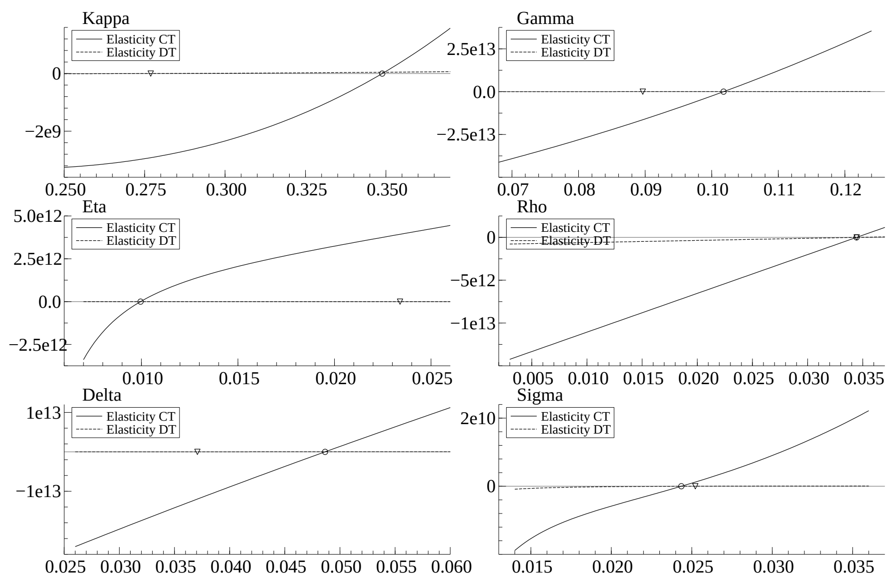
Figure D2: Simulation Study – Objective Function with Elasticity

The figure reports the elasticity of the objective function to the parameter values for both the continuous time and discrete time models. For each model, one parameter is varied over the range of the horizontal axis while the other parameters are fixed at the estimated values for each method. Both the criterion function and elasticity (percentage change of objective function divided by percentage change of parameter value) are reported, where the objective function is -1 times the inner product of the elements of $M_T(\phi)$. The figure is an illustration for 1 of the 1,000 replications in the simulation study, with generated 25 years of data from the underlying data generating process (DGP) in the case of monthly data. Panel A plots the criterion function for both the continuous (solid line) and discrete time (dashed line) models, where the dot (reverse triangle) is the estimated parameter value. Panel B plots the elasticity for both the continuous (solid line) and discrete time (dashed line) models, where the dot (reverse triangle) is the estimated parameter value.

(A) Objective Function – Continuous Time (CT) and Discrete Time (DT)



(B) Objective Function Elasticity – Continuous Time (CT) and Discrete Time (DT)



E Appendix Tables

Table E1: Simulation Study – Sensitivity to DGP values

The table reports output of a simulation study into the sensitivity of the Table 1 monthly results for the OLS (Panel A), FGLS-SUR-IV (Panel B), GMM (Panel C), and MEF (Panel D) methods to the parameter settings used in the Data Generating Process (DGP). In each Panel, the top row reports the baseline DGP settings and the second row the estimates obtained for these settings (these are the estimates of Table 1). Then we vary one parameter at a time and consider two settings for each, one value lower than the one used in the baseline setting, and one value higher than that of the baseline DGP setting (while keeping all other parameters at the baseline settings). In all cases, for 1,000 replications, we generate 25 years of data from the underlying DGP and apply our estimation strategy. We show the median estimate, and provide the interquartile range below it.

Panel A: Parameter Estimates from Simulation Study – OLS Sensitivity to DGP values

	κ	γ	η	ρ	δ	σ
Baseline DGP settings	0.200	0.100	0.010	0.030	0.050	0.020
OLS for Baseline DGP	0.349	0.201	0.083	0.080	0.050	0.317
	0.286	0.036	0.035	0.015		0.040
DGP with $\kappa = 0.1$	0.272	0.200	0.070	0.079	0.050	0.313
	0.251	0.055	0.038	0.019		0.058
DGP with $\kappa = 0.5$	0.628	0.201	0.112	0.081	0.050	0.319
	0.348	0.018	0.033	0.009		0.024
DGP with $\gamma = 0.05$	0.325	0.087	0.033	0.049	0.050	0.193
	0.283	0.036	0.023	0.014		0.063
DGP with $\gamma = 0.2$	0.354	0.412	0.174	0.135	0.050	0.460
	0.287	0.067	0.070	0.033		0.067
DGP with $\eta = 0.005$	0.354	0.206	0.087	0.083	0.050	0.325
	0.286	0.034	0.035	0.017		0.047
DGP with $\eta = 0.05$	0.175	0.054	0.002	0.032	0.050	0.000
	0.310	0.221	0.040	0.009		0.094
DGP with $\rho = 0.01$	0.349	0.201	0.083	0.060	0.050	0.317
	0.286	0.036	0.035	0.015		0.040
DGP with $\rho = 0.1$	0.349	0.201	0.083	0.150	0.050	0.317
	0.286	0.036	0.035	0.015		0.040
DGP with $\delta = 0.01$	0.349	0.201	0.083	0.080	0.010	0.317
	0.286	0.036	0.035	0.015		0.040
DGP with $\delta = 0.1$	0.349	0.201	0.083	0.080	0.100	0.317
	0.286	0.036	0.035	0.015		0.040
DGP with $\sigma = 0.01$	0.348	0.201	0.083	0.080	0.050	0.317
	0.284	0.034	0.035	0.012		0.037
DGP with $\sigma = 0.05$	0.351	0.200	0.083	0.080	0.050	0.319
	0.289	0.045	0.038	0.026		0.061

Table E1, Panel B: Parameter Estimates from Simulation Study – FGLS-SUR-IV Sensitivity to DGP values						
	κ	γ	η	ρ	δ	σ
Baseline DGP settings	0.200	0.100	0.010	0.030	0.050	0.020
FGLS-SUR-IV for Baseline DGP	0.299	0.101	0.008	0.030	0.050	0.000
	0.134	0.013	0.004	0.006		<0.001
DGP with $\kappa = 0.1$	0.263	0.101	0.009	0.030	0.050	0.000
	0.146	0.022	0.004	0.006		<0.001
DGP with $\kappa = 0.5$	0.403	0.100	0.006	0.030	0.050	0.000
	0.150	0.006	0.003	0.006		<0.001
DGP with $\gamma = 0.05$	0.404	0.051	0.008	0.030	0.050	0.000
	0.214	0.012	0.003	0.006		<0.001
DGP with $\gamma = 0.2$	0.220	0.201	0.006	0.030	0.050	0.000
	0.117	0.014	0.006	0.006		0.020
DGP with $\eta = 0.005$	0.220	0.100	0.002	0.030	0.050	0.012
	0.117	0.007	0.005	0.006		0.020
DGP with $\eta = 0.05$	0.541	0.151	0.029	0.030	0.050	0.000
	0.305	0.045	0.038	0.006		<0.001
DGP with $\rho = 0.01$	0.299	0.101	0.008	0.010	0.050	0.000
	0.134	0.013	0.004	0.006		<0.001
DGP with $\rho = 0.1$	0.299	0.101	0.008	0.100	0.050	0.000
	0.134	0.013	0.004	0.006		<0.001
DGP with $\delta = 0.01$	0.299	0.101	0.008	0.030	0.010	0.000
	0.134	0.013	0.004	0.006		<0.001
DGP with $\delta = 0.1$	0.299	0.101	0.008	0.030	0.100	0.000
	0.134	0.013	0.004	0.006		<0.001
DGP with $\sigma = 0.01$	0.298	0.101	0.009	0.030	0.050	0.000
	0.134	0.013	0.003	0.003		<0.001
DGP with $\sigma = 0.05$	0.301	0.100	0.000	0.030	0.050	0.037
	0.136	0.013	0.007	0.014		0.020

Table E1, Panel C: Parameter Estimates from Simulation Study – GMM Sensitivity to DGP values

	κ	γ	η	ρ	δ	σ
Baseline DGP settings	0.200	0.100	0.010	0.030	0.050	0.020
GMM for Baseline DGP	0.345	0.100	0.010	0.031	0.050	0.027
	0.345	0.014	0.001	0.007		0.047
DGP with $\kappa = 0.1$	0.265	0.100	0.010	0.031	0.050	0.026
	0.312	0.024	0.001	0.007		0.049
DGP with $\kappa = 0.5$	0.594	0.101	0.010	0.031	0.050	0.025
	0.449	0.006	0.001	0.007		0.044
DGP with $\gamma = 0.05$	0.335	0.053	0.010	0.031	0.050	0.033
	0.327	0.013	0.001	0.007		0.062
DGP with $\gamma = 0.2$	0.290	0.200	0.010	0.031	0.050	0.033
	0.312	0.015	0.001	0.007		0.048
DGP with $\eta = 0.005$	0.330	0.100	0.005	0.030	0.050	0.022
	0.319	0.007	$<$0.001	0.006		0.038
DGP with $\eta = 0.05$	0.352	0.197	0.050	0.042	0.050	0.157
	0.351	0.554	0.050	0.265		0.678
DGP with $\rho = 0.01$	0.344	0.100	0.010	0.011	0.050	0.027
	0.348	0.014	0.001	0.007		0.047
DGP with $\rho = 0.1$	0.345	0.100	0.010	0.101	0.050	0.027
	0.345	0.014	0.001	0.007		0.047
DGP with $\delta = 0.01$	0.344	0.100	0.010	0.031	0.010	0.027
	0.345	0.014	0.001	0.007		0.047
DGP with $\delta = 0.1$	0.344	0.100	0.010	0.031	0.100	0.027
	0.346	0.014	0.001	0.007		0.047
DGP with $\sigma = 0.01$	0.352	0.101	0.010	0.031	0.050	0.020
	0.361	0.014	0.001	0.003		0.044
DGP with $\sigma = 0.05$	0.344	0.100	0.010	0.031	0.050	0.052
	0.325	0.015	0.001	0.017		0.032

Table E1, Panel D: Parameter Estimates from Simulation Study – MEF Sensitivity to DGP values						
	κ	γ	η	ρ	δ	σ
Baseline DGP settings	0.200	0.100	0.010	0.030	0.050	0.020
MEF for Baseline DGP	0.354	0.099	0.010	0.030	0.050	0.023
	0.284	0.013	0.001	0.006	0.002	0.005
DGP with $\kappa = 0.1$	0.212	0.099	0.010	0.030	0.050	0.020
	0.239	0.026	0.001	0.006	0.002	0.003
DGP with $\kappa = 0.5$	0.624	0.100	0.010	0.030	0.050	0.021
	0.350	0.005	0.001	0.006	0.001	0.003
DGP with $\gamma = 0.05$	0.393	0.050	0.010	0.030	0.050	0.021
	0.324	0.014	0.001	0.006	0.002	0.006
DGP with $\gamma = 0.2$	0.356	0.199	0.010	0.030	0.050	0.021
	0.282	0.013	0.001	0.006	0.002	0.004
DGP with $\eta = 0.005$	0.351	0.100	0.005	0.030	0.050	0.021
	0.286	0.006	0.000	0.006	0.001	0.003
DGP with $\eta = 0.05$	0.578	0.143	0.049	0.030	0.051	0.023
	0.723	0.049	0.006	0.007	0.037	0.130
DGP with $\rho = 0.01$	0.355	0.099	0.010	0.010	0.050	0.022
	0.283	0.013	0.001	0.006	0.002	0.004
DGP with $\rho = 0.1$	0.356	0.099	0.010	0.100	0.050	0.019
	0.284	0.013	0.001	0.006	0.002	0.001
DGP with $\delta = 0.01$	0.356	0.099	0.010	0.030	0.010	0.023
	0.282	0.013	0.001	0.005	0.002	0.005
DGP with $\delta = 0.1$	0.357	0.099	0.010	0.030	0.100	0.020
	0.288	0.013	0.001	0.006	0.002	0.004
DGP with $\sigma = 0.01$	0.355	0.099	0.010	0.030	0.050	0.011
	0.294	0.013	0.001	0.003	0.002	0.001
DGP with $\sigma = 0.05$	0.356	0.099	0.010	0.030	0.049	0.054
	0.282	0.013	0.001	0.014	0.002	0.010

Table E2: Simulation Study – Monthly Data with Bias Correction

The table reports output of a simulation study of the accuracy of the structural model parameters estimated using the MEF approach to the AK-Vasicek model, where bias correction methods are applied. We apply the formulas from Yu (2012, eq. (17)), Tang and Chen (2009, Theorem 3.1.1), and two bootstrap methods inspired by Tang and Chen (2009, Section 4), where we bias correct based on both the mean and median simulated bias. For 1,000 replications, we generate 25 years of data from the underlying continuous time data generating process (DGP) and apply our estimation strategy. We show the median estimate, and provide the interquartile range below it.

Parameter Estimates from Simulation Study – Monthly Data with Bias Correction Methods

			Yu (2012)	Tang and Chen (2009)	Bootstrapped	
	DGP	MEF	(17)	Theorem 3.1.1	Mean	Median
κ	0.200	0.355	0.278	0.192	0.150	0.204
		0.285	0.280	0.283	0.306	0.313
γ	0.100	0.099			0.099	0.099
		0.013			0.013	0.013
η	0.010	0.010		0.010	0.010	0.010
		0.001		0.001	0.001	0.001
ρ	0.030	0.030			0.030	0.030
		0.006			0.006	0.005
δ	0.050	0.050			0.050	0.050
		0.002			0.002	0.002
σ	0.020	0.023			0.020	0.022
		0.005			0.006	0.005

Table E3: Simulation Study – Variance Terms and Five Conditional Moment Restrictions

The table reports output of a simulation study of the incorporation of additional moments for the OLS, FGLS-SUR-IV, GMM, and MEF approaches to the AK-Vasicek model. For 1,000 replications, we generate 25 years of data from the underlying data generating process (DGP) and apply our estimation strategy. We show the median estimate, and provide the interquartile range below it.

Parameter Estimates from Simulation Study – Monthly & Quarterly Data									
		Monthly Data				Quarterly Data			
	DGP	OLS	FGLS-SUR-IV	GMM	MEF	OLS	FGLS-SUR-IV	GMM	MEF
κ	0.200	0.168	0.299	0.311	0.285	0.171	0.227	0.177	0.244
		0.141	0.134	0.348	0.425	0.155	0.119	0.265	0.422
γ	0.100	0.100	0.100	0.102	0.100	0.100	0.101	0.110	0.100
		0.015	0.013	0.014	0.015	0.015	0.013	0.059	0.019
η	0.010	0.010	0.009	0.010	0.010	0.010	0.010	0.010	0.010
		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
ρ	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030
		0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
δ	0.050	0.050	0.050	0.051	0.050	0.050	0.051	0.054	0.050
		<0.001	0.001	0.002	0.002	<0.001	0.001	0.067	0.004
σ	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.019	0.020
		0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002

Table E4: Simulation Study – Sensitivity of Regression-Based Methods to δ_0 and σ_0

The table reports output of a simulation study of the sensitivity of the Table 1 monthly results for the regression-based OLS and FGLS-SUR-IV methods to the δ_0 and σ_0 settings. For δ_0, we consider the values 0.01, 0.05 (base), and 0.10, and for σ_0, we consider the values 0.01, 0.02 (base), and 0.05. We set the restricted value for δ equal to δ_0 for internal consistency. Panel A reports the performance of the OLS method, and Panel B that of the FGLS-SUR-IV method. Each panel consists of nine columns, where each column represents a δ_0 and σ_0 combination. For 1,000 replications, we generate 25 years of data from the underlying data generating process (DGP) and apply our estimation strategy. We show the median estimate, and provide the interquartile range below it.

Panel A: Parameter Estimates from Simulation Study – OLS Sensitivity to δ_0 and σ_0

		$\delta_0 = 0.01$			$\delta_0 = 0.05$			$\delta_0 = 0.10$		
	DGP	$\sigma_0 = 0.01$	$\sigma_0 = 0.02$	$\sigma_0 = 0.05$	$\sigma_0 = 0.01$	$\sigma_0 = 0.02$	$\sigma_0 = 0.05$	$\sigma_0 = 0.01$	$\sigma_0 = 0.02$	$\sigma_0 = 0.05$
κ	0.200	0.292	0.295	0.305	0.349	0.349	0.350	0.354	0.354	0.354
		0.281	0.282	0.286	0.286	0.286	0.286	0.281	0.281	0.282
γ	0.100	0.108	0.109	0.114	0.201	0.201	0.206	0.312	0.313	0.317
		0.034	0.034	0.033	0.036	0.036	0.036	0.053	0.054	0.054
η	0.010	0.039	0.040	0.042	0.083	0.083	0.085	0.131	0.131	0.133
		0.026	0.026	0.026	0.035	0.035	0.036	0.051	0.051	0.052
ρ	0.030	0.054	0.055	0.057	0.080	0.080	0.083	0.111	0.112	0.114
		0.012	0.012	0.012	0.015	0.015	0.015	0.024	0.024	0.025
δ	0.050	0.010	0.010	0.010	0.050	0.050	0.050	0.100	0.100	0.100
σ	0.020	0.222	0.224	0.234	0.316	0.317	0.324	0.401	0.402	0.408
		0.051	0.050	0.047	0.041	0.040	0.040	0.057	0.057	0.058

Panel B: Parameter Estimates from Simulation Study – FGLS-SUR-IV Sensitivity to δ_0 and σ_0

		$\delta_0 = 0.01$			$\delta_0 = 0.05$			$\delta_0 = 0.10$		
	DGP	$\sigma_0 = 0.01$	$\sigma_0 = 0.02$	$\sigma_0 = 0.05$	$\sigma_0 = 0.01$	$\sigma_0 = 0.02$	$\sigma_0 = 0.05$	$\sigma_0 = 0.01$	$\sigma_0 = 0.02$	$\sigma_0 = 0.05$
κ	0.200	0.196	0.201	0.223	0.299	0.299	0.297	0.109	0.108	0.104
		0.210	0.211	0.201	0.134	0.134	0.134	0.058	0.058	0.058
γ	0.100	0.067	0.068	0.071	0.101	0.101	0.102	0.161	0.161	0.163
		0.041	0.039	0.031	0.013	0.013	0.013	0.019	0.019	0.020
η	0.010	0.018	0.019	0.021	0.009	0.008	0.000	0.000	0.000	0.000
		0.026	0.024	0.020	0.003	0.004	0.004	$<$0.001	$<$0.001	$<$0.001
ρ	0.030	0.038	0.038	0.039	0.030	0.030	0.031	0.032	0.032	0.033
		0.008	0.008	0.008	0.006	0.006	0.006	0.006	0.006	0.007
δ	0.050	0.010	0.010	0.010	0.050	0.050	0.050	0.100	0.100	0.100
σ	0.020	0.132	0.133	0.140	0.000	0.000	0.036	0.058	0.061	0.077
		0.033	0.032	0.028	$<$0.001	$<$0.001	0.018	0.042	0.040	0.033

Table E5: Estimates – Variance Terms and 5 Moment Conditions

The table reports estimates for the structural model parameters estimated using OLS, FGLS-SUR-IV, GMM, and MEF approaches for the AK-Vasicek model. For OLS and FGLS-SUR-IV, we use the variance terms for the consumption and interest rate equation, and for GMM and MEF, we use five conditional moment restrictions. We run the estimation for monthly data (where production is measured by IP) and quarterly data (production measured by GDP). The sample runs from January, 1982, through December, 2012. Asymptotic t-statistics are given below the estimates.

Parameter Estimates from Empirical Data								
	Monthly Data				Quarterly Data			
	OLS	FGLS-SUR-IV	GMM	MEF	OLS	FGLS-SUR-IV	GMM	MEF
κ	0.096	0.083	0.030	0.069	0.114	0.065	0.045	0.048
	0.436	0.270	0.185	0.679	1.064	2.083	0.769	0.697
γ	0.101	0.101	0.045	0.108	0.134	0.130	0.089	0.098
	4.002	1.671	0.186	0.602	2.715	2.329	2.054	0.924
η	0.018	0.018	0.005	0.007	0.028	0.019	0.000	0.007
	1.284	0.444	0.669	0.051	0.693	0.608	$<$0.001	0.097
ρ	0.015	0.015	0.006	0.004	0.022	0.021	0.009	0.020
	0.441	1.139	0.153	0.089	0.672	0.957	0.444	0.538
δ	0.098	0.106	0.050	0.081	0.128	0.153	0.050	0.040
	1.298	1.443		0.243	0.732	0.682		0.173
σ	0.018	0.018	0.014	0.018	0.018	0.017	0.000	0.019
	0.065	0.082	0.994	0.008	0.078	0.003	$<$0.001	0.010

Table E6: Simulation Study – Iterated MEF Approach

The table reports output of a simulation study of the accuracy of the structural model parameters estimated using the iterated MEF approaches for the AK-Vasicek model. For 100 replications, we generate 25 years of data from the underlying data generating process (DGP) and apply our estimation strategy. We report estimates using the MEF approach with both three and five moment conditions for the regular MEF and iterated approach, for two iterations. We show the median estimate, and provide the interquartile range below it.

Parameter Estimates from Simulation Study – Iterated MEF 3 and 5 Moments									
		Monthly Data				Quarterly Data			
		3 Conditions		5 Conditions		3 Conditions		5 Conditions	
	DGP	MEF	two-step MEF	MEF	two-step MEF	MEF	two-step MEF	MEF	two-step MEF
κ	0.200	0.348	0.239	0.288	0.200	0.353	0.316	0.241	0.209
		0.309	0.202	0.480	0.204	0.310	0.339	0.366	0.129
γ	0.100	0.100	0.109	0.101	0.104	0.099	0.130	0.100	0.107
		0.012	0.048	0.014	0.017	0.013	0.074	0.021	0.021
η	0.010	0.010	0.001	0.010	0.010	0.010	0.000	0.010	0.010
		0.001	0.002	0.001	0.000	0.002	0.001	0.001	0.001
ρ	0.030	0.030	0.032	0.030	0.031	0.030	0.032	0.030	0.030
		0.005	0.010	0.006	0.007	0.006	0.016	0.005	0.012
δ	0.050	0.050	0.059	0.050	0.051	0.050	0.072	0.050	0.055
		0.002	0.045	0.002	0.008	0.003	0.063	0.005	0.019
σ	0.020	0.022	0.025	0.020	0.020	0.024	0.030	0.020	0.020
		0.005	0.018	0.001	0.001	0.011	0.030	0.002	0.003

References

An, S., and F. Schorfheide (2007): "Bayesian Analysis of DSGE models," *Econometric Rev.*, 26(2-4), 113–172.

Canova, F., and L. Sala (2009): "Back to square one: Identification issues in DSGE models," *J. Monet. Econ.*, 56, 431–449.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Rev. Econ. Stud.*, 74, 1059–1087.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of Understanding VARs," *Amer. Econ. Rev.*, 97(3), 1021–1026.

Ross, S. M. (2014): *Introduction to Probability Models.* Elsevier, 11th ed., Amsterdam.

Tang, C. Y., and S. X. Chen (2009): "Parameter estimation and bias correction for diffusion processes," *J. Econometrics*, 149, 65–81.

Vasicek, O. (1977): "An Equilibrium Characterization of the Term Structure," *J. Finan. Econ.*, 5, 177–188.

Wang, X., P. C. B. Phillips, and J. Yu (2011): "Bias in estimating multivariate and univariate diffusions," *J. Econometrics*, 161, 228–245.

Yu, J. (2012): "Bias in the estimation of the mean reversion parameter in continuous time models," *J. Econometrics*, 169, 114–122.