

Generalized Autoregressive Method of Moments

Revision: July 2018

Drew Creal¹

Siem Jan Koopman^{2,3}

André Lucas^{2,3}

Marcin Zamojski^{2,3}

¹ The University of Chicago Booth School of Business, United States

² Faculty of Economics and Business Administration, VU University Amsterdam

³ Tinbergen Institute, the Netherlands

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Generalized autoregressive Method of Moments*

Drew Creal¹, Siem Jan Koopman^{2,3,4} André Lucas^{2,3} and Marcin Zamojski⁵

¹ University of Notre Dame, Notre Dame IN 46556, USA

² Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

³ Tinbergen Institute, 1082 MS Amsterdam, The Netherlands

⁴ CREATES, Aarhus University, 8210 Aarhus, Denmark

⁵ Centre for Finance, University of Gothenburg, 405 30 Gothenburg, Sweden

ABSTRACT

We extend the generalized method of moments to a setting where a subset of the parameters may vary over time with unknown dynamics. We approximate the true unknown dynamics by an updating scheme that is driven by the influence function of the conditional criterion function at time t . The updates ensure a local improvement of the conditional criterion function at each time in expectation. In our framework, time-varying parameters are a function of past data; it leads to a computationally efficient method since it does not require simulation-based methods for estimation. The approach can be applied to a wide range of moment conditions that are used in economics and finance. We provide an illustration for a capital asset pricing model with time-varying risk aversion.

Keywords: dynamic models, time-varying parameters, non-linearity, score-driven dynamics.

JEL classification codes: C10, C22, C32, C51.

*Corresponding Author: S.J. Koopman, Vrije Universiteit Amsterdam, School of Business and Economics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: s.j.koopman@vu.nl. Phone: +31205986019. Fax: +31205986020. Emails: D. Creal dcreal@nd.edu S.J. Koopman s.j.koopman@vu.nl A. Lucas alucas@vu.nl M. Zamojski marcin.zamojski@gu.se. Creal gratefully acknowledges financial support from the William Ladany Faculty Scholar Fund at the University of Chicago Booth School of Business. Lucas and Zamojski are grateful to the Dutch National Science Foundation (NWO, grant VICI453-09-005) for financial support. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78) at Aarhus University, Denmark, and funded by the Danish National Research Foundation; as well as support from the European Union Seventh Framework Programme (FP7-SSH/2007-2013, grant agreement 320270 - SYRTO).

1. Introduction

The Generalized Method of Moments (GMM) framework of Hansen (1982) has had a major impact on the development of time series econometrics in the last three decades and continues to be widely used in economics and finance. GMM provides a unified framework for inference, while only requiring the specification of a vector of moment conditions that are often derived from economic theory. The approach is able to handle partially-defined dynamic models for economic data. GMM also allows the use of lagged variables as instruments which is particularly appealing as the dynamic relationships between economic variables are typically complex. For a review of the literature, see Hansen and West (2002) and Kuersteiner (2012).

When GMM is applied to time series data, an important step of the analysis is accounting for possible instabilities in estimated parameters. Much research has focused on the case of identifying structural breaks in parameters; see, for example, Andrews (1993). However, in empirical work it is often found that additional model flexibility is required to match the key features in economic and financial time series, such as frequent changes in conditional means, volatilities, or correlations. The importance of capturing these features has been highlighted by Shephard (2005), Hamilton (2010), and Fernández-Villaverde and Rubio-Ramírez (2013), among others. In this context, time-varying parameter models have become a popular alternative to structural breaks and regime switching models. So far, most contributions on time-varying parameters are in the context of likelihood-based inference (using both frequentist and Bayesian methods) while much less research in this direction is conducted within the GMM framework. In our study we extend the GMM framework by allowing some parameters to vary over time in an integrated and general fashion, as well as in a computationally appealing way.

In our extension of GMM, we estimate the true time-varying parameter using an autoregressive updating scheme driven by the *influence function* of the conditional GMM criterion function. This new parameter updating scheme adjusts the time-varying parameter estimates in a (local) steepest descent direction based on the information contained in the new incoming observations and their discordance with local moment conditions (at time t). In this way, the recursive updating scheme accounts for the sequential process of data becoming available over time: a real-time mechanism that is typical for the analysis of time series data. We refer to our method as the Generalized autoregressive Method of Moments (GaMM). The GaMM updating scheme approximates the true time-varying parameter by a function of only past data. Our approach is observation-driven in the

classification of [Cox \(1981\)](#) and as such it is computationally straightforward. We argue that GaMM is a highly effective method, especially in settings where other techniques are difficult to implement or are computationally prohibitive, and where no alternative methods are readily available. In addition to extending the flexibility of GMM by accommodating time-varying parameters, GaMM retains all appealing features of GMM.

The influence function that we use to drive the parameter dynamics is well-known and has been applied in many areas of statistics. For instance, influence functions are used to measure changes in estimators of static parameters when adding new observations and have implications for robustness, optimality, and efficiency of estimators; see [Hampel et al. \(2011\)](#). Interestingly, one can show that standard Gauss-Newton-Raphson optimization algorithms adjust static parameters based on the average value of their influence functions; see the in-depth discussions in [Cox and Hinkley \(1979\)](#). We develop this interpretation of the influence function further towards the time-varying parameter case. The intuition is clear: large values of the influence function for a series of observations over time suggest that local adjustments of the parameters can result in a better local fit of the moment conditions. GaMM incorporates this information into the updating scheme of the time-varying parameter values by using the influence function as the driving mechanism. Although our focus in this paper is on the GMM framework, this driving mechanism can be applied to any criterion function for which an influence function exists and can be computed. This includes other M-estimators such as maximum likelihood or quantile estimators; see [Patton et al. \(2017\)](#).

The updating scheme based on the influence function is not arbitrary. We show that a GaMM update leads to local improvements *in expectation* of the conditional GMM criterion function. This result holds under different forms of mis-specification of the model, the underlying parameter dynamics, or the moment conditions. [Blasques et al. \(2015\)](#) establish similar, although more restrictive optimality results for the special case of maximum likelihood estimation (MLE) and the generalized autoregressive score dynamics (GAS) of [Creal et al. \(2011, 2013\)](#) and [Harvey \(2013\)](#). We generalize these optimality results to a general setting of local criterion functions and parameter update steps; for an alternative approach in the context of GMM, see [Gospodinov and Otsu \(2012\)](#).

Although GaMM is designed to extend the standard GMM framework for time series, it also generalizes other observation-driven approaches proposed in the literature. In particular, GaMM encompasses the generalized autoregressive score dynamics (GAS) of [Creal et al. \(2011, 2013\)](#) and [Harvey \(2013\)](#), which itself has generalized many earlier models. Hence GaMM nests many

popular econometric models which include the GARCH model of Engle (1982) and Bollerslev (1986) and many of its variations, the autoregressive conditional duration (ACD) model of Engle and Russell (1998), as well as many new models for time-varying parameters under fat-tails and mixed observation densities; see the references in Creal et al. (2011) as well as those in Harvey and Luati (2014), Creal et al. (2014), Patton et al. (2017), Opschoor et al. (2018), Lucas et al. (2018), and Oh and Patton (2018). For this strand of literature, the influence function perspective provides a clear motivation for the type of scaling (inverse conditional Fisher information matrix) used in GAS. Previously, the scaling has been dealt with in a rather ad-hoc manner, as in Creal et al. (2011, 2013), or it has not been implemented at all, as in Oh and Patton (2018).

In our empirical study, we apply GaMM to a simple consumption-based asset pricing model with myopic expectation formation and an unstable risk aversion parameter. As in the seminal work of Hansen and Singleton (1982), we use a constant relative risk aversion (CRRA) utility function. Many contributions in the literature have argued that this simple model fails at explaining the cross-sectional variation in stock returns: time series estimates of static risk aversion in this model are often too high compared to risk aversion estimates obtained from experimental data. This phenomenon is referred to as the equity premium puzzle; see, for example, Savov (2011), Mehra and Prescott (1985), Chen and Ludvigson (2009), Lettau and Ludvigson (2009) or Ludvigson (2011). Typically, the estimates of the CRRA risk aversion coefficient in the standard model appear sensitive to the choice of the sample period, initial parameter values, and the choice of instruments. More recently, Malmendier and Nagel (2011); Cohn et al. (2015); Guiso et al. (2018) provide empirical and experimental evidence for instability of risk-aversion among households, finance professionals, and investors respectively. Here, we consider a GaMM specification of the CRRA utility function with a time-varying risk aversion parameter. A possible alternative considered in the literature is a specification of utility function that includes (consumption) habit formation; see, for example, Constantinides (1990), Campbell and Cochrane (1999) or Ludvigson (2011). This also leads to a time-varying relative risk aversion that is a direct function of consumption dynamics. In our GaMM framework, by contrast, we do not need to explicitly impose what drives the changes in risk aversion. The path of the time-varying risk aversion parameter is recovered from the ever changing relationship between past returns and past consumption growth.

Using a static CRRA specification for U.S. quarterly consumption and stock return data between 1947 and 2015, we confirm the findings in Savov (2011) and Lettau and Ludvigson (2009): the GMM estimates of the relative risk aversion and the discount rate do not have any economic

intuition. Risk aversion is estimated implausibly high (> 100), and the discount rate exceeds one. When considering a time-varying risk aversion coefficients in our GaMM framework, much more reasonable average values of the risk aversion coefficient (between 1 and 3) are obtained. The results are qualitatively robust against the use of other proxies for consumption; see, for example, Kroencke (2017). Overall, we conclude that within the CRRA setting the data strongly favour a specification with a time-varying relative risk aversion coefficient. The empirics thus provide an effective illustration of the potential of GaMM as a tool for empirically investigating possible parameter instability in a GMM framework.

The remainder of this paper is organized as follows. In Section 2 we lead with a motivating example as well as an intuitive explanation of the main ideas. Section 3 develops the methodology in its generic form. Section 4 presents our empirical dynamic non-linear asset pricing application for U.S. consumption and stock return data. Section 5 concludes.

2. A regression example

To provide intuition for our approach and a better understanding of the challenge of the problem, we first consider a linear regression model for a dependent variable y_t with a $k \times 1$ vector of regressors \mathbf{x}_t and with a disturbance ε_t that has mean zero and constant variance. The k regressors are possibly endogenous such that we have moment conditions $E[\mathbf{x}_t \varepsilon_t] \neq 0$. The variables y_t and \mathbf{x}_t are time series variables that are subject to different dynamic processes and are possibly part of a larger dynamic system that can be subject to shocks and breaks. The assumption of a static regression coefficient may therefore not be very realistic and hence we consider a time-varying process for the regression parameter β_t . We thus face the major challenge of treating endogenous regressors and time-varying parameters simultaneously.

Under endogeneity but with a static parameter vector, we could base our analysis on a $K \times 1$ vector of instrumental variables \mathbf{z}_t , with $K \geq k$, and apply GMM using moment conditions $E[\mathbf{z}_t \varepsilon_t] = 0$. Conversely, under a time-varying regression parameter but with exogenous regressors, we could treat the time-varying parameters by a range of methods including the celebrated Kalman filter. The simultaneous treatment of both endogenous regressors and time-varying parameters, however, is not straightforward: GMM cannot easily be generalized to the time-varying parameter setting, while the Kalman filter provides biased filtered estimates when regressors are endogenous.

Our framework starts from the *conditional* moment conditions $E_{t-1}[\mathbf{z}_t \varepsilon_t] = 0$, where $E_{t-1}[\cdot]$ refers

to conditioning on past information. These conditional moment conditions at time t may contain information about the direction in which to change the time-varying parameters to obtain a better local fit of the model. For instance, if specific elements of $E_{t-1}[\mathbf{z}_t \varepsilon_t]$ are persistently positive over some period of time, the fit to the moment conditions can be improved substantially by adjusting the corresponding elements in $\boldsymbol{\beta}_t$ temporarily upwards (or downwards) using a steepest descent direction based on the time t conditional GMM criterion function.

We formalize this mechanism through an autoregressive updating scheme, which we refer to as the Generalized autoregressive Method of Moments (GaMM). This leads to a mechanism that lets the parameters change gradually and continuously over time. This updating scheme exploits the possible persistent deviations in the empirical realizations of time t moment conditions. In essence, GaMM incorporates new information into the parameter dynamics by removing time series predictability in the conditional moment conditions.

In our GaMM framework, we assume that the true dynamics of $\boldsymbol{\beta}_t$ are unknown. We filter these dynamics from the data using functions of past observations only. Hence our approach belongs to the class of observation-driven models in the classification of [Cox \(1981\)](#) and can therefore exploit major computational and statistical advantages. To define the updating steps for $\boldsymbol{\beta}_t$, we consider the conditional moment conditions at time t and the corresponding local objective function

$$E_{t-1}[\mathbf{z}_t(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}_t)]^\top \boldsymbol{\Omega}_{t-1} E_{t-1}[\mathbf{z}_t(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}_t)], \quad (1)$$

where $\boldsymbol{\Omega}_{t-1}$ is a weighting matrix, which for the sake of this illustration we set equal to the unit matrix, $\boldsymbol{\Omega}_{t-1} = \mathbf{I}$. The update of $\boldsymbol{\beta}_t$ aims to improve this local objective function given the incoming observation $(y_t, \mathbf{x}_t^\top, \mathbf{z}_t^\top)^\top$. The update is based on the influence function of the GMM estimator. The influence function can be interpreted as a ‘derivative’ of the time t conditional GMM objective function, where the derivative is taken with respect to the time-varying parameter $\boldsymbol{\beta}_t$ in the direction of the new data point $(y_t, \mathbf{x}_t^\top, \mathbf{z}_t^\top)$. Intuitively, this yields a steepest descent improvement to the time t conditional GMM criterion function based on the most recent data.

Motivating example. *In summary, consider a dependent variable $y_t \in \mathbb{R}$, a vector of regressors $\mathbf{x}_t \in \mathbb{R}^{k \times 1}$, and a time-varying parameter vector $\boldsymbol{\beta}_t$ that constitute a time-varying regression model $y_t = \mathbf{x}_t^\top \boldsymbol{\beta}_t + \varepsilon_t$ where disturbance ε_t has mean zero and constant variance. We assume that some of the regressors are endogenous, such that $E_{t-1}[\mathbf{x}_t \varepsilon_t] \neq 0$. We adopt the local objective function in eq. (1) for a vector of valid instruments $\mathbf{z}_t \in \mathbb{R}^{K \times 1}$ with $K \geq k$ and $E_{t-1}[\mathbf{z}_t \varepsilon_t] = 0$.*

The steepest descent update step of the time-varying parameter as based on the influence function in this case reduces to

$$\mathbf{s}_t = C_t(\mathbf{x}, \mathbf{z}) \mathbf{z}_t (y_t - \mathbf{x}_t^\top \boldsymbol{\beta}_t), \quad (2)$$

where $C_t(\mathbf{x}, \mathbf{z})$ is a scaling matrix that depends on the second moments of \mathbf{x}_t and \mathbf{z}_t ; see Section 3 for more details. We incorporate the ‘steepest descent’ step \mathbf{s}_t in an autoregressive updating scheme for $\boldsymbol{\beta}_t$ as follows:

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\omega} \cdot (\mathbf{I} - \mathbf{B}) + \mathbf{B} \boldsymbol{\beta}_t + \mathbf{A} \mathbf{s}_t, \quad (3)$$

where $\boldsymbol{\omega}$, \mathbf{A} , and \mathbf{B} are static vectors and matrices that are unknown and need to be estimated. For reasons of parsimony, \mathbf{A} and \mathbf{B} can be restricted to diagonal matrices or even to scalars. We refer to the updating scheme eq. (3) as GaMM(1, 1). The “(1, 1)” refers to the use of one lag for both the time-varying parameter $\boldsymbol{\beta}_t$ and the updating step \mathbf{s}_t . We can easily generalize the updating step by including more lags for $\boldsymbol{\beta}_t$, \mathbf{s}_t , or both, or by introducing more complex dynamic schemes, including fractionally integrated lag polynomials.

The GaMM updating scheme in eq. (2) and (3) is intuitive as we can interpret s_t as the contribution of the time t observation to the GMM estimator. In Section 3.2 we show that the GaMM update always ensures an *expected* local improvement to the conditional GMM objective function. This convenient property generalizes the notion of optimal score-based steps in a Kullback-Leibler framework as studied by Blasques et al. (2015). Effectively, the arguments developed below in our study of GaMM extend to a much wider class of conditional objective functions for which an influence function can be defined.

A final step to operationalize GaMM is an estimation approach for the new static parameters $\boldsymbol{\omega}$, \mathbf{A} and \mathbf{B} that govern the dynamics of $\boldsymbol{\beta}_t$ in eq. (3). This requires a new set of moment conditions. For example, in the simple regression case of $k = K = 1$, we use the instrumented unconditional moment conditions

$$\mathbb{E} [\mathbf{z}_t \cdot (y_t - \mathbf{x}_t^\top \boldsymbol{\beta}_t) \cdot (1, \boldsymbol{\beta}_{t-1}, \mathbf{s}_{t-1})^\top] = 0, \quad (4)$$

where $(1, \boldsymbol{\beta}_{t-1}, \mathbf{s}_{t-1})^\top$ is a vector of instruments. We can recognize the instruments as the partial derivatives of $\boldsymbol{\beta}_t$ with respect to $\boldsymbol{\omega}$, \mathbf{B} , and \mathbf{A} . We show in Section 3.3 that these proposed

instruments can be regarded as efficient approximations to the optimal instruments for ω , A , and B , respectively. Furthermore, the instruments are numerically easy to compute as they are obtained as a direct by-product of the recursion in eq. (3). Considering the unconditional moment conditions in eq. (4) in more detail, we see that they exploit the persistence in the time series of s_t by including terms such as $z_t(y_t - x_t^\top \beta_t)s_{t-1} = s_t s_{t-1}$. Since eq. (4) imposes that the unconditional expectation of this cross-product must be zero, the dynamic scheme in eq. (3) removes as much autocorrelation in the time t moment condition $z_t(y_t - x_t^\top \beta_t)$ as possible.

Summarizing, our GaMM framework effectively consists of three main steps. First, we allow some of the parameters in the standard GMM set-up to be time-varying with dynamics of unknown form. This is effectively as simple as providing a time index for some of the parameters, though the choice of which parameters can vary over time should naturally be rooted in economic theory. Second, we approximate these unknown dynamics by a non-linear autoregressive scheme driven only by past observations. Our key innovation is the use of the influence function of the time t conditional GMM estimator to drive the parameter dynamics. Such an updating scheme ensures the local improvement of the conditional GMM criterion in expectation. Intuitively, it provides steepest descent improvements of the local GMM criterion function at time t . Third, to estimate the new static parameters in the non-linear autoregressive GaMM updating scheme, we introduce feasible instruments and a corresponding unconditional GMM criterion. The main illustration used in this section of the standard linear regression model with endogenous regressors is carried through into Section 3 when setting up GaMM for general non-linear moment conditions, establishing its optimality properties, and its asymptotic properties.

3. The Generalized autoregressive Method of Moments

In this section, we consider the generalized autoregressive method of moments (GaMM) in its general form, we provide theoretical arguments justifying the GaMM update steps as expected improvements of the local GMM objective function at time t and we illustrate the results for the linear regression model with endogenous regressors as discussed in the previous section.

Motivating example (ctd). *In our main example of a linear regression model with time-varying regression parameters and endogenous regressors, define $V_{xx,t} = E_{t-1}[x_t x_t^\top]$, $V_{zz,t} = E_{t-1}[z_t z_t^\top]$, $V_{zx,t} = E_{t-1}[z_t x_t^\top]$ and $V_{xz,t} = V_{zx,t}^\top$, and assume that all three matrices have full column rank. These conditions, together with $E_{t-1}[z_t \varepsilon_t] = 0$, ensure valid instruments.*

3.1. General form of GaMM dynamics

Consider the conditional moment condition at time t ,

$$\mathbb{E}_{t-1} [\mathbf{g}_t (\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)] = 0, \quad (5)$$

where $\mathbf{g}_t : \mathbb{R}^m \times \mathcal{F} \times \Theta_1 \rightarrow \mathbb{R}^K$ is a vector-valued function, $\mathbf{w}_t \in \mathbb{R}^m$ is the data vector, and $(\mathbf{f}_t, \boldsymbol{\theta}_1)$ is a parameter vector that is split into a time-varying parameter $\mathbf{f}_t \in \mathcal{F}$ and a static parameter $\boldsymbol{\theta}_1 \in \Theta_1$. We approximate the unknown dynamics of \mathbf{f}_t by observation-driven dynamics, i.e., we write \mathbf{f}_t as a function of past observations $\mathbf{w}_{t-1}, \mathbf{w}_{t-2}, \dots$, only. The moment function $g_t(\cdot)$ can be non-linear in \mathbf{f}_t .

To introduce the observation-driven GaMM dynamics for \mathbf{f}_t , consider the time t conditional GMM criterion function

$$\mathbb{E}_{t-1} [\mathbf{g}_t (\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)]^\top \boldsymbol{\Omega}_{t-1} \mathbb{E}_{t-1} [\mathbf{g}_t (\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)], \quad (6)$$

where $\boldsymbol{\Omega}_{t-1}$ is a positive definite weighting matrix known at time $t - 1$. The expectations in eq. (6) are computed under the time $t - 1$ conditional measure $F_{\mathbf{w}}$ of \mathbf{w}_t .

Motivating example (ctd). *In the regression model with time-varying regression parameter and endogeneity, we have $\mathbf{f}_t = \beta_t$ and $\mathbf{w}_t^\top = (y_t, \mathbf{x}_t^\top, \mathbf{z}_t^\top)$. The function defining the moment condition is given by $\mathbf{g}_t (\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1) = z_t(y_t - \mathbf{x}_t^\top \mathbf{f}_t)$, which in this case does not depend on $\boldsymbol{\theta}_1$.*

To propagate \mathbf{f}_t forward to \mathbf{f}_{t+1} given the realization of \mathbf{w}_t , we take a scaled steepest descent step using an appropriate concept of a derivative of eq. (6). A similar approach in a likelihood based framework was introduced by [Creal et al. \(2011, 2013\)](#) and [Harvey \(2013\)](#) using a standard (Newtonian) derivative of the log-density at time t . The GMM criterion eq. (6) is different from the likelihood framework in that we minimize *the square* of an average (or expectation) rather than the average (or expectation) of a log-density. We therefore use a corresponding, more appropriate derivative concept, namely the Gâteaux derivative. The Gâteaux derivative leads directly to the well-known concept of the influence function, which plays a central role in many parts of statistics; see, for example, [Hampel et al. \(2011\)](#). The influence function or Gâteaux derivative in the direction of the new data point \mathbf{w}_t accounts in a natural way for new information about the position of the time-varying parameter \mathbf{f}_t compared to the incoming data \mathbf{w}_t . We prove later that the influence

function steps lead to *expected* local improvements of eq. (6), even if the model is mis-specified and the parameter dynamics do not coincide with those of the data generating process.

To define the influence function, consider a contaminated measure $F_{\mathbf{w}}^\epsilon = (1 - \epsilon)F_{\mathbf{w}} + \epsilon \delta_{w_t}$, where δ_{w_t} is the Dirac measure that puts unit mass on the realized value w_t . Overweighting the new observation \mathbf{w}_t by using $F_{\mathbf{w}}^\epsilon$ rather than $F_{\mathbf{w}}$ allows us to capture the information in the new observation in an appropriate way when updating \mathbf{f}_t to \mathbf{f}_{t+1} .

The first order condition with respect to \mathbf{f}_t corresponding to eq. (6), evaluated at the measure $F_{\mathbf{w}}^\epsilon$, is given by

$$(\bar{\mathbf{G}}_t^\epsilon)^\top \Omega_{t-1} E_{t-1}^\epsilon [\mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)] = 0, \quad (7)$$

with

$$\mathbf{G}_{\mathbf{f},t} = \frac{\partial \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)}{\partial \mathbf{f}_t^\top}, \quad \bar{\mathbf{G}}_{\mathbf{f},t}^\epsilon = E_{t-1}^\epsilon [\mathbf{G}_{\mathbf{f},t}], \quad \bar{\mathbf{G}}_{\mathbf{f},t} = \bar{\mathbf{G}}_{\mathbf{f},t}^0,$$

and where E_{t-1}^ϵ and $E_{t-1} \equiv E_{t-1}^0$ denote the expectation operators with respect to $F_{\mathbf{w}}^\epsilon$ and $F_{\mathbf{w}}^0 \equiv F_{\mathbf{w}}$, respectively. Note that the \mathbf{f}_t that solves eq. (7) is a function of $F_{\mathbf{w}}^\epsilon$, i.e., $\mathbf{f}_t = \mathbf{f}_t(F_{\mathbf{w}}^\epsilon)$. The following result is well-known.

Proposition 1 (see, e.g., Hampel et al., 2011 or Ronchetti and Trojani, 2001). *Assuming eq. (5), and taking derivatives of eq. (7) with respect to ϵ and evaluating at $\epsilon = 0$, we obtain the influence function*

$$\mathbf{s}_t = \left. \frac{d\mathbf{f}_t(F_{\mathbf{w}}^\epsilon)}{d\epsilon} \right|_{\epsilon=0} = - \left((\bar{\mathbf{G}}_{\mathbf{f},t})^\top \Omega_{t-1} \bar{\mathbf{G}}_{\mathbf{f},t} \right)^{-1} (\bar{\mathbf{G}}_{\mathbf{f},t})^\top \Omega_{t-1} \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1), \quad (8)$$

where the inverse can be replaced by a pseudo-inverse if $\bar{\mathbf{G}}_{\mathbf{f},t}^\top \Omega_{t-1} \bar{\mathbf{G}}_{\mathbf{f},t}$ is rank deficient.

Motivating example (ctd). For our example, we have $\bar{\mathbf{G}}_{\mathbf{f},t} = E_{t-1}[-\mathbf{z}_t \mathbf{x}_t^\top] = -\mathbf{V}_{zx,t}$. If we set Ω_{t-1} to its typical, ‘optimal’ choice $\Omega_{t-1} = \mathbf{V}_{zz,t}^{-1}$, we obtain

$$\mathbf{s}_t = (\mathbf{V}_{xz,t} \mathbf{V}_{zz,t}^{-1} \mathbf{V}_{zx,t})^{-1} \mathbf{V}_{xz,t} \mathbf{V}_{zz,t}^{-1} \mathbf{z}_t (y_t - \mathbf{x}_t^\top \mathbf{f}_t),$$

The step \mathbf{s}_t reflects the projection of the \mathbf{x}_t variables on the space spanned by the instruments as is typical for instrumental variables estimation.

The step \mathbf{s}_t can be directly used in a random walk scheme for updating \mathbf{f}_t , that is $\mathbf{f}_{t+1} = \mathbf{f}_t + \mathbf{a} \mathbf{s}_t$ for some step length $\mathbf{a} > 0$. More generally, we propose to base the dynamics on the autoregressive updating scheme

$$\mathbf{f}_{t+1} = \omega + \sum_{j=1}^p \mathbf{B}_j (\mathbf{f}_{t+1-j} - \omega) + \sum_{i=1}^q \mathbf{A}_i \mathbf{s}_{t+1-i}, \quad (9)$$

where $\omega = \omega(\boldsymbol{\theta}_2)$, $\mathbf{B}_j = \mathbf{B}_j(\boldsymbol{\theta}_2)$, and $\mathbf{A}_i = \mathbf{A}_i(\boldsymbol{\theta}_2)$ are appropriately sized vectors and matrices that depend on a static parameter vector $\boldsymbol{\theta}_2$, and where we extend the parameter vector to $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)$. We refer to the updating scheme in eq. (9) as GaMM(p, q) dynamics. In particular, in this paper we work with the GaMM(1,1) specification with $p = q = 1$, $\mathbf{A}_1 = \mathbf{A}$, and $\mathbf{B}_1 = \mathbf{B}$, as given by

$$\mathbf{f}_{t+1} = (\mathbf{I} - \mathbf{B}) \boldsymbol{\omega} + \mathbf{B} \mathbf{f}_t + \mathbf{A} \mathbf{s}_t. \quad (10)$$

The random walk scheme clearly is a special case of eq. (10) with $\boldsymbol{\omega} = \mathbf{0}$, $\mathbf{A}_1 = \mathbf{a}\mathbf{I}$, $\mathbf{B}_1 = \mathbf{I}$.

If eq. (5) holds, $\{\mathbf{s}_t\}_{t \in \mathbb{Z}}$ forms a martingale difference sequence with $E_{t-1}[\mathbf{s}_t] = 0$, where the expectation is with respect to the probability measure F_w . Due to the properties of martingale differences, the unconditional mean of \mathbf{f}_t in eq. (9) is $E[\mathbf{f}_t] = \boldsymbol{\omega}$ if \mathbf{f}_t is weakly stationary and eq. (5) holds. To initialize the recursion in eq. (9), we can therefore set $\mathbf{f}_t = \omega$ for $t = 0, -1, \dots, -p + 1$ and $\mathbf{s}_t = 0$ for $t = 0, -1, \dots, -q + 1$. Alternatively, these initial values can be estimated either jointly with the static parameters in $\boldsymbol{\theta}$ or with a static model based on a small estimation window at the beginning of the sample period. Note, however, that such estimates for the initial conditions are not consistent.

Interestingly, the steps based on the influence function simplify to the generalized autoregressive score-driven (GAS) approach of [Creal et al. \(2011, 2013\)](#) and [Harvey \(2013\)](#) if the moment condition is the first order condition of the log-likelihood criterion. To see this, let $\mathbf{g}(\mathbf{w}_t; \mathbf{f}_t; \boldsymbol{\theta}_1) = \partial \log p(\mathbf{w}_t | \mathbf{f}_t; \boldsymbol{\theta}_1) / \partial \mathbf{f}_t$, where $p(\mathbf{w}_t | \mathbf{f}_t; \boldsymbol{\theta}_1)$ is the time t conditional predictive density of \mathbf{w}_t . Filling this out in eq. (8), we obtain

$$\mathbf{s}_t = \bar{\mathbf{G}}_{\mathbf{f}, t}^{-1} \cdot (\partial \log p(\mathbf{w}_t | \mathbf{f}_t; \boldsymbol{\theta}_1) / \partial \mathbf{f}_t),$$

which precisely coincides with the scaled score step of [Creal et al. \(2013\)](#) using the inverse conditional Fisher information matrix to scale the score. Whereas this scaling was introduced rather ad-hoc in [Creal et al. \(2013\)](#), it follows without any further ad-hoc assumptions from the direct

definition of the influence function in our GaMM framework. The property that GAS is a special case of GaMM mirrors the well-known fact that maximum likelihood is a special case of GMM.¹ As the score-driven approach of [Creal et al. \(2013\)](#) in turn encompasses many well-known time-varying parameter models from the literature such as GARCH, MEM, ACD, and many other models, the GaMM approach covers an even wider range of empirically relevant models and settings.

3.2. GaMM steps as expected improvements to moment conditions

In this sub-section, we provide a further theoretical motivation for the use of the steps s_t in eq. (8). In particular, we show that the influence function steps s_t ensure local improvements to the time t conditional objective function eq. (6) *in expectation*. The result extends optimality results for score-based steps in a likelihood framework as derived by [Blasques et al. \(2015\)](#). [Blasques et al. \(2015\)](#) show for the generalized autoregressive score model of [Creal et al. \(2011, 2013\)](#) that time-varying parameter updates based on the score of the time t log-likelihood function improve the local Kullback–Leibler divergence between the true, unknown data density and the model density in areas where the true density has much probability mass. In addition, they show that any observation-driven update with similar optimality properties needs to be proportional to the score. Their results hold under very general forms of mis-specification of the postulated statistical model. Score-based updates apparently locally improve the Kullback–Leibler divergence between the true data generating process and the statistical model, even if the latter is mis-specified.

The earlier results of [Blasques et al.](#) were presented in the context of information-theoretic optimality and Kullback–Leibler divergence. This fits nicely with the original score-based approach of [Creal et al. \(2013\)](#), which explicitly relies on the likelihood framework and on the conditional observation density. In the context of GMM, however, optimality centers around the quadratic objective function of the moment conditions as in eq. (6). As a result, we need to adapt the concepts and results in [Blasques et al. \(2015\)](#) accordingly. Similarly to [Blasques et al.](#), we can argue that our results below hold irrespective of whether or not the model is mis-specified. In particular, we formulate conditions that ensure local improvements in expectation even if the model is incorrect, moment conditions are violated, or parameter dynamics are not correctly specified.

Define a short-hand notation for the time t GMM objective function in eq. (6) evaluated at some

¹The updates in [Creal et al. \(2011, 2013\)](#) and [Harvey \(2013\)](#) can also be directly derived as the influence function of the maximum (log-)likelihood criterion.

\mathbf{f} as

$$\mathcal{C}(t, \mathbf{f}) = \left(\int \mathbf{g}_t(\mathbf{w}; \mathbf{f}, \boldsymbol{\theta}_1) dF_{\mathbf{w}}(\mathbf{w}) \right)^T \boldsymbol{\Omega}_{t-1} \left(\int \mathbf{g}_t(\mathbf{w}; \mathbf{f}, \boldsymbol{\theta}_1) dF_{\mathbf{w}}(\mathbf{w}) \right), \quad (11)$$

where as before $F_{\mathbf{w}}$ denotes the distribution of \mathbf{w}_t conditional on all past observations. Note we do not require the moment conditions to be correctly specified, that is, we do not require $E_{t-1}[\mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta})] = 0$. Even if the model and the moment conditions were correctly specified, we should not expect that our initial guess of \mathbf{f}_1 would be correct and, at least during the start-up phase of the model, we expect that the moment conditions are not met at the postulated values for \mathbf{f}_t . It is then important to know whether the estimated time-varying parameter \mathbf{f}_t moves in the ‘correct’ direction.

We make the following assumptions.

Assumption A.

A.1 *The moment function $\mathbf{g}_t(\mathbf{w}; \mathbf{f}, \boldsymbol{\theta})$ is twice continuously differentiable in \mathbf{f} with measurable derivatives (in \mathbf{w}_t).*

A.2 *The weight matrix $\boldsymbol{\Omega}_{t-1}$ is positive definite and finite. The derivative matrix $\bar{\mathbf{G}}_t$ is finite and has full column rank.*

A.3 $E_{t-1}[\mathbf{g}_{j,t}(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)^4] < \infty$ with $\mathbf{g}_{j,t}(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)$ the j th element of $\mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)$.

A.4 $\sup_{\mathbf{f}} \|\int \partial \mathbf{G}_{\mathbf{f}, t}(\mathbf{w}; \mathbf{f}, \boldsymbol{\theta}_1) / \partial \mathbf{f} dF_{\mathbf{w}}(\mathbf{w})\| \leq C < \infty$ for every $\boldsymbol{\theta}_1$.

In the Appendix, we prove the following result.

Proposition 2. *Given Assumption A, there exists a GaMM(1,1) update with $\boldsymbol{\omega} = 0$, $\mathbf{B} = \mathbf{I}$, and $\mathbf{A} = \mathbf{a} \mathbf{I}$, such that*

$$E_{t-1} [\mathcal{C}(t, \mathbf{f}_t) - \mathcal{C}(t, \mathbf{f}_{t+1})] \geq 0, \quad (12)$$

where \mathbf{f}_{t+1} is given by the recursion in equation eq. (10).

The main insight of Proposition 2 is that a GaMM(1,1) update with $\boldsymbol{\omega} = 0$, $\mathbf{B} = \mathbf{I}$, and $\mathbf{A} = \mathbf{a} \mathbf{I}$ for small positive \mathbf{a} results in a non-increase of the local objective function *in expectation*. Informally stated, GaMM steps *on average* improve the local GMM criterion function at time t , or at least, do not increase it. This provides a strong basis for the influence function based steps as introduced

in Section 3.1. Proposition 2 holds without specifying the true conditional distribution of the data F_w and without assuming that the model or the moment conditions are correctly specified. Along the same lines as in Blasques et al. (2015) for a likelihood based framework, one can further extend these results to establish non-local improvement properties and improvement properties for the more general autoregressive scheme in eq. (9) if one is willing to impose additional conditions on the data generating process.

The main assumptions needed for Proposition 2 are the smoothness of the moment conditions, the finiteness of both the weight matrix Ω_{t-1} and of the expected moment derivatives when evaluated at f_t , the existence of sufficient moments of the score steps, and a uniform boundedness condition on the moment derivatives matrix. The latter two may possibly be relaxed further, but the sufficient conditions in Assumption A are enough for our current purposes.

Motivating example (ctd). For our example, we have $\bar{G}_{f,t} = E_{t-1}[-z_t x_t^\top] = -V_{zx,t}$, which does not depend on f_t , such that $\partial \bar{G}_{f,t} / \partial f = 0$ and A.4 is directly satisfied. It is also easy to show that A.1 and A.2 are met. What remains are the moment assumptions on $z_{i,t}^4 y_t^4$ and $z_{i,t}^4 x_{j,t}^4$, where $z_{i,t}$ and $x_{j,t}$ are the i -th and j -th element of z_t and x_t , respectively. The Appendix contains an elaborate simulation experiment for this model to show that GaMM(1,1) dynamics indeed succeed in tracking the unknown parameter dynamics of f_t , even if the model is severely mis-specified.

3.3. Estimation of static parameters and choice of instruments

The estimation of the new static parameters ω , A , and B in θ_2 requires the formulation of new, additional moment conditions. To solve this issue and to estimate the static parameters in θ , we augment the conditional moment conditions in eq. (5) by a matrix of instruments $Z_t \in \mathbb{R}^{L \times K}$ to arrive at the unconditional moment conditions

$$E[Z_t g_t(w_t; f_t, \theta_1)] = 0, \quad (13)$$

and the corresponding unconditional GMM criterion function

$$\min_{\theta \in \Theta} \tilde{g}_T^\top \tilde{\Omega}_T \tilde{g}_T, \quad \tilde{g}_T = \frac{1}{T} \sum_{t=1}^T Z_t g_t(w_t; f_t, \theta_1), \quad (14)$$

where $\tilde{\Omega}_T$ is a positive definite matrix. Note that f_t itself also depends on θ , both via the recurrence relation in eq. (10) and via the specification of the influence function step s_t . We make this explicit

by writing $\mathbf{f}_t = \mathbf{f}_t(\mathbf{w}_{t-1}; \mathbf{f}_{t-1}, \boldsymbol{\theta})$ where needed. As usual, we can start estimating $\boldsymbol{\theta}$ by setting $\tilde{\boldsymbol{\Omega}}_T = \mathbf{I}$ in a first-step estimation and minimizing eq. (14). In a second step, we can then replace $\tilde{\boldsymbol{\Omega}}_T$ by a long run variance estimate of $\text{Var} [T^{1/2} \tilde{\mathbf{g}}_T]$ obtained from the first step; see Hansen (1982).

As instruments we propose $\mathbf{Z}_t = (1, \mathbf{f}_{t-1}^\top, \mathbf{s}_{t-1}^\top)^\top \otimes \mathbf{I}$, such that

$$\mathbb{E} [\mathbf{Z}_t \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)] = \mathbb{E} [(1, \mathbf{f}_{t-1}^\top, \mathbf{s}_{t-1}^\top)^\top \otimes \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)] = 0, \quad (15)$$

where \otimes denotes the Kronecker product. These instruments are easy to compute and come as a by-product of the construction of \mathbf{f}_t itself. One can easily recognize \mathbf{Z}_t as the partial derivative of \mathbf{f}_t with respect to $(\mathbf{I} - \mathbf{B})\boldsymbol{\omega}$, \mathbf{B} , and \mathbf{A} , respectively. The instrumented moment condition $\mathbb{E}[1 \otimes \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)] = 0$ imposes that the conditional moment condition should also hold unconditionally. The other two instrumented moment conditions $\mathbb{E}[\mathbf{s}_{t-1} \otimes \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)] = 0$ and $\mathbb{E}[\mathbf{f}_{t-1} \otimes \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)] = 0$ exploit the autocorrelation in the sample values of the moment conditions \mathbf{g}_t . In particular, $\mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1) \otimes \mathbf{s}_{t-1}$ holds the cross products of the moment conditions \mathbf{g}_t and its lags \mathbf{g}_{t-1} via the lagged scores \mathbf{s}_{t-1} . If there is any first-order autocorrelation in \mathbf{g}_t , the unconditional moment conditions in eq. (14) remove this autocorrelation as much as possible by exploiting the GaMM dynamics and adjusting \mathbf{A} and \mathbf{B} accordingly.

One way to motivate eq. (15) is as an approximate solution to the optimal instrument problem; see, e.g. Davidson and MacKinnon (1993) for a textbook treatment on optimal instruments. Deriving optimal instruments in a general setting is typically non-trivial. To facilitate the exposition, we make a simplifying assumption that the moment conditions are correctly specified in the sense that for all $t < s$ we have

$$\mathbb{E} [\mathbf{g}_s(\mathbf{w}_s; \mathbf{f}_s, \boldsymbol{\theta}_1) \mathbf{g}_t^\top(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)] = 0. \quad (16)$$

We concentrate on the optimal instruments for the GaMM dynamics, i.e., for $\boldsymbol{\theta}_2$. Following Davidson and MacKinnon (1993), these are given by

$$\begin{aligned} \mathbf{Z}_{\theta_2, t}^\top &= \boldsymbol{\Omega}_{t-1} \mathbb{E}_{t-1} \left[\frac{d\mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)}{d\boldsymbol{\theta}_2^\top} \right] \\ &= \boldsymbol{\Omega}_{t-1} \mathbb{E}_{t-1} \left[\frac{\partial \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)}{\partial \mathbf{f}_t^\top} \frac{d\mathbf{f}_t}{d\boldsymbol{\theta}_2^\top} \right] = \boldsymbol{\Omega}_{t-1} \bar{\mathbf{G}}_{f,t}^\top \frac{d\mathbf{f}_t}{d\boldsymbol{\theta}_2^\top}, \end{aligned} \quad (17)$$

where the last equality follows from the fact that \mathbf{f}_t depends on past data only due to the observation

driven nature of GaMM dynamics. The total derivative $d\mathbf{f}_t/d\boldsymbol{\theta}_2^\top$ is the key part for the *dynamic* \mathbf{f}_t in eq. (17). Note that the product $\boldsymbol{\Omega}_{t-1}\bar{\mathbf{G}}_{f,t}^\top$ would also be there even if \mathbf{f}_t were a static rather than a dynamic parameter. The total derivative in the instruments in eq. (17) and the partial derivative proposed as our instruments in eq. (15) are clearly different, as the latter does not account for the fact that \mathbf{f}_t and \mathbf{s}_t on the right-hand side of eq. (8) depend on $\boldsymbol{\theta}$. However, we have the following result.

Proposition 3. *Given a GaMM(1,1) update with $\boldsymbol{\omega} = 0$, $\mathbf{B} = (1 - \mathbf{a})\mathbf{I}$, and $\mathbf{A} = \mathbf{a}\mathbf{I}$, then for small $\mathbf{a} > 0$*

$$\frac{d\mathbf{f}_t}{d\boldsymbol{\theta}_2^\top} = \mathbf{I} \otimes \begin{pmatrix} 1 & \overrightarrow{\mathbf{f}_{t-1}}^\top & \overrightarrow{\mathbf{s}_{t-1}}^\top \end{pmatrix} + \mathcal{O}(\mathbf{a}), \quad (18)$$

where $\overrightarrow{\mathbf{s}_{t-1}}$ and $\overrightarrow{\mathbf{f}_{t-1}}$ are exponentially weighted moving averages of \mathbf{s}_{t-i} and \mathbf{f}_{t-i} for $i \geq 1$, respectively, and where $\mathcal{O}(\mathbf{a})$ vanishes at order \mathbf{a} for $\mathbf{a} \rightarrow 0$.

The difference between using the partial rather than the total derivative of \mathbf{f}_t with respect to $\boldsymbol{\theta}_2$ as instruments thus amounts to using unsmoothed versus exponentially smoothed versions of $(\mathbf{s}_{t-1}, \mathbf{f}_{t-1})$. The unsmoothed instruments provide a quick approximation to the dominant part of the optimal instruments for $\boldsymbol{\theta}_2$ without the need to keep track of the additional recursion for $d\mathbf{f}_t/d\boldsymbol{\theta}$ as spelled out in Appendix A (page S.4). In fact, the instruments are obtained as a direct by-product of the GaMM recursion. The simulations in the subsequent sections reveal that the instruments \mathbf{Z}_t work well empirically: the resulting estimates of \mathbf{f}_t track the true parameter dynamics well, even in cases where the model is mis-specified.

Motivating example (ctd). *For our example with a time-varying regression parameter in an instrumental variables setting, we refer to the simulations in Appendix B. These show that static parameter estimates and estimated paths of the time-varying parameter are accurate. The bias in the estimated paths of \mathbf{f}_t compared to for instance a standard linear Gaussian state-space model are virtually zero for the GaMM approach, and only reveal a modest increase in variability of the estimates.*

3.4. Asymptotic distribution theory

To derive the asymptotic properties of $\hat{\boldsymbol{\theta}}$ such as consistency and asymptotic normality, we can largely build on the standard framework for GMM estimators as originally developed by Hansen

(1982). We need to differentiate between two sequences \mathbf{f}_t . The first sequence is defined for $t = 1, 2, \dots$ and initialized at $\bar{\mathbf{f}}_1$ and follows the (initialized) stochastic recurrence in eq. (10). We write this sequence as $\{\mathbf{f}_t(\boldsymbol{\theta}; \bar{\mathbf{f}}_1)\}_{t \in \mathbb{N}}$. The second sequence is defined for $t = \dots, -1, 0, 1, \dots$ and is the stationary and ergodic solution of eq. (10). We write this second sequence as $\{\mathbf{f}_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$. We make the following assumptions.

Assumption B.

- B.1 For a neighborhood Θ_0 of $\boldsymbol{\theta}_0$, the process $\{\mathbf{f}_t(\boldsymbol{\theta}; \bar{\mathbf{f}}_1)\}_{t \in \mathbb{N}}$ initialized at $\bar{\mathbf{f}}_1$ converges uniformly and exponentially fast almost surely to a unique stationary and ergodic sequence $\{\mathbf{f}_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$, i.e., $c_1^t \sup_{\boldsymbol{\theta} \in \Theta_0} \|\mathbf{f}_t(\boldsymbol{\theta}; \bar{\mathbf{f}}_1) - \mathbf{f}_t(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0$ for some $c_1 > 0$ and for every $\bar{\mathbf{f}}_1 \in \mathcal{F}$. A similar requirement holds for the first two derivatives of $\{\mathbf{f}_t(\boldsymbol{\theta}; \bar{\mathbf{f}}_1)\}_{t \in \mathbb{N}}$ with respect to $\boldsymbol{\theta}$.
- B.2 The process $\{\mathbf{Z}_t(\mathbf{f}_t(\boldsymbol{\theta})) \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_1)\}_{t \in \mathbb{Z}}$ is stationary and ergodic with finite first moment, with $\mathbf{Z}_t(\mathbf{f}_t(\boldsymbol{\theta}))^\top = (1, \mathbf{f}_{t-1}(\boldsymbol{\theta})^\top, \mathbf{s}_{t-1}(\mathbf{f}_{t-1}(\boldsymbol{\theta}); \boldsymbol{\theta})^\top)$ and \mathbf{g}_t twice continuously differentiable in $\mathbf{f}_t(\boldsymbol{\theta})$ and measurable in \mathbf{w}_t for every $\boldsymbol{\theta}$. Moreover, the scaling matrix $\tilde{\Omega}_T \xrightarrow{a.s.} \tilde{\Omega}$, where $\tilde{\Omega}$ is positive definite.
- B.3 The GMM objective function (14) and its limiting objective function

$$\mathbb{E}[\mathbf{Z}_t \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_1)]^\top \tilde{\Omega} \mathbb{E}[\mathbf{Z}_t \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_1)]$$

are twice continuously differentiable with respect to $\boldsymbol{\theta}$. The limiting objective function has a unique minimum at $\boldsymbol{\theta}_0^* \in \text{int}(\Theta)$, where $\text{int}(\Theta)$ denotes the interior of the compact parameter space Θ .

- B.4 The gradient $d\tilde{\mathbf{g}}_T/d\boldsymbol{\theta}^\top$ and Hessian $d\tilde{\mathbf{g}}_T/(d\boldsymbol{\theta}^\top \otimes d\boldsymbol{\theta}^\top)$ from eq. (14) evaluated at $\boldsymbol{\theta}_0^*$ and at the stationary and ergodic sequence $\{\mathbf{f}_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ satisfy a uniform law of large numbers (ULLN) with full column rank limits $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{H}}$, respectively.
- B.5 The standardized unconditional moment condition $\sqrt{T}\tilde{\mathbf{g}}_T$ of eq. (14) evaluated at $\boldsymbol{\theta}_0^*$ and at the stationary and ergodic sequence $\{\mathbf{f}_t(\boldsymbol{\theta}_0^*)\}_{t \in \mathbb{Z}}$ satisfies a central limit theorem (CLT) and weakly converges to a multivariate normal with mean zero and covariance matrix $\tilde{\mathbf{V}}$.

Assumptions B.2–B.5 are standard high-level conditions for the identification, consistency, and asymptotic normality of GMM estimators for stationary and ergodic data sequences; see for instance Hansen (1982), Davidson and MacKinnon (1993), or White (1996), including a further breakdown to some more restrictive low-level conditions. This underlines once more that the GaMM set-up

seamlessly matches the familiar standard GMM framework. Note that conditions [B.2–B.3](#) allow for $E[\mathbf{Z}_t \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_1)] \neq 0$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0^*$. As shown in [Section 3.2](#), the local improvement property of the GaMM steps continues to hold under this type of mis-specification, though convergence is no longer to the ‘true’ parameter, but to a pseudo-true parameter $\boldsymbol{\theta}_0^*$ that minimizes the limiting GMM objective function.

Condition [B.1](#) is familiar in the time-varying parameter context and relates to the concept of continuous filter invertibility. The importance of invertibility has been stressed before by for instance [Straumann and Mikosch \(2006\)](#) and [Wintenberger \(2013\)](#). It ensures that the filter converges pathwise to a unique stationary and ergodic solution, irrespective of the starting condition $\bar{\mathbf{f}}_1$. The exponential rate of convergence is used to establish the asymptotic normality of $\hat{\boldsymbol{\theta}}$ later on. It is not needed to establish consistency. For many models, Assumption [B.1](#) takes a familiar form. For instance, for GARCH models the conditions collapse to the familiar convergence conditions formed by “Nelson’s triangle”; see [Nelson \(1990, 1991\)](#). For the exponential GARCH (EGARCH) model, primitive conditions for Assumption [B.1](#) to hold are formulated in [Straumann and Mikosch \(2006\)](#) and [Wintenberger \(2013\)](#), while primitive conditions for the highly non-linear class of univariate generalized autoregressive score models are given by [Blasques et al. \(2014\)](#). Also for our leading example, condition [B.1](#) takes a simple form.

Motivating example (ctd). *For our motivating example with a time-varying regression parameter in an instrumental variables setting with stationary and ergodic regressors, we have*

$$\mathbf{s}_t = (\mathbf{V}_{xz}\mathbf{V}_{zz}^{-1}\mathbf{V}_{zx})^{-1}\mathbf{V}_{xz}\mathbf{V}_{zz}^{-1}\mathbf{z}_t(y_t - \mathbf{x}_t^\top \mathbf{f}_t).$$

The expression for \mathbf{f}_{t+1} is then linear in \mathbf{f}_t with a random matrix coefficient:

$$\begin{aligned} \mathbf{f}_{t+1} = & (\mathbf{I} - \mathbf{B})\boldsymbol{\omega} + \\ & \mathbf{A}(\mathbf{V}_{xz}\mathbf{V}_{zz}^{-1}\mathbf{V}_{zx})^{-1}\mathbf{V}_{xz}\mathbf{V}_{zz}^{-1}\mathbf{z}_t y_t + \\ & \left(\mathbf{B} - \mathbf{A}(\mathbf{V}_{xz}\mathbf{V}_{zz}^{-1}\mathbf{V}_{zx})^{-1}\mathbf{V}_{xz}\mathbf{V}_{zz}^{-1}\mathbf{z}_t \mathbf{x}_t^\top \right) \mathbf{f}_t. \end{aligned}$$

For exponential convergence, besides requiring $E[\mathbf{z}_t y_t]$ to be finite, we need the largest Lyapunov exponent of this stochastic recurrence equation to be negative. This is implied by requiring the

spectral radius of $\mathbf{B} - \mathbf{A}$ to be smaller than 1, as

$$\mathbb{E} \left[\mathbf{B} - \mathbf{A} \left(\mathbf{V}_{xz} \mathbf{V}_{zz}^{-1} \mathbf{V}_{zx} \right)^{-1} \mathbf{V}_{xz} \mathbf{V}_{zz}^{-1} \mathbf{z}_t \mathbf{x}_t^\top \right] = \mathbf{B} - \mathbf{A}.$$

For instance, for the time-varying mean model with $\mathbf{x}_t = \mathbf{z}_t = 1$, this collapses to $|\mathbf{B} - \mathbf{A}| < 1$.

We now have the following result.

Proposition 4. *Under Assumption B, the GMM estimator for the static parameter vector $\boldsymbol{\theta}$, including the parameters $\boldsymbol{\omega}$, \mathbf{A} and \mathbf{B} specifying the GaMM dynamics, is consistent and asymptotically normal, i.e.,*

$$T^{1/2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \tilde{\mathbf{K}}^{-1} \tilde{\mathbf{L}} \tilde{\mathbf{K}}^{-1}), \quad (19)$$

where

$$\tilde{\mathbf{K}} = (\tilde{\mathbf{G}}^\top \tilde{\mathbf{V}}^{-1} \tilde{\mathbf{G}})^{-1} \quad \tilde{\mathbf{L}} = \tilde{\mathbf{G}}^\top \tilde{\boldsymbol{\Omega}} \tilde{\mathbf{V}} \tilde{\boldsymbol{\Omega}} \tilde{\mathbf{G}},$$

where $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{V}}$ are defined in Assumption B. The efficient weighting matrix is $\tilde{\boldsymbol{\Omega}} = \tilde{\mathbf{V}}^{-1}$, in which case the asymptotic covariance matrix collapses to $(\tilde{\mathbf{G}}^\top \tilde{\mathbf{V}}^{-1} \tilde{\mathbf{G}})^{-1}$.

The key of the proof of this result as shown in the Appendix is that we can replace the non-stationary initialized filtered series $\{\mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)\}_{t=1}^\infty$ in the expressions for the GMM estimator by its stationary and ergodic limit $\{\mathbf{f}_t(\boldsymbol{\theta})\}_{t=-\infty}^\infty$. After that, the proof follows the standard line of argument for GMM estimators for stationary and ergodic data sequences.

3.5. Extension: exploiting scores of static parameters

In this section, we present a possible extension of the GaMM(1,1) approach that exploits the misfit in the time t moment conditions both with respect to the dynamic parameter \mathbf{f}_t and with respect to the static parameter $\boldsymbol{\theta}_1$. The advantage of this extension is that it can capture more empirical features in the data. For instance, when modelling volatility of stock returns, it can give rise to the so called leverage type effect, i.e., an asymmetric response of volatility dynamics to positive and negative returns.

As a first step, we note that a static parameter is a special case of GaMM(1,1). This follows immediately if we set $\mathbf{A} = 0$ and $\mathbf{B} = 0$. To exploit this feature further, we treat the static

parameter $\boldsymbol{\theta}_1$ and the time-varying parameter \mathbf{f}_t jointly as $\tilde{\mathbf{f}}_t^\top \equiv (\mathbf{f}_t^\top, \boldsymbol{\theta}_1^\top)$. The GaMM(1,1) specification for $\tilde{\mathbf{f}}_t$ is given by

$$\tilde{\mathbf{f}}_{t+1} = \tilde{\boldsymbol{\omega}} + \tilde{\mathbf{B}} \left(\tilde{\mathbf{f}}_t - \tilde{\boldsymbol{\omega}} \right) + \tilde{\mathbf{A}} \tilde{\mathbf{s}}_t, \quad (20)$$

where

$$\tilde{\mathbf{B}} = \begin{bmatrix} \mathbf{B} & \mathbf{C}^B \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \tilde{\boldsymbol{\omega}} = \begin{bmatrix} \boldsymbol{\omega} \\ \boldsymbol{\theta}_1 \end{bmatrix}, \quad \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{C}^A \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (21)$$

with $\tilde{\mathbf{s}}_t$ containing the influence function with respect to parameter vector $\tilde{\mathbf{f}}_t$ rather than \mathbf{f}_t only.

So far, we considered the matrices \mathbf{C}^A and \mathbf{C}^B to be zero. These restrictions, however, can be relaxed. By allowing for non-zero matrices \mathbf{C}^A and \mathbf{C}^B , the time-varying parameter \mathbf{f}_t also responds to the influence function of the static parameter $\boldsymbol{\theta}_1$. This may sometimes be useful. For example, when analyzing a time series of financial returns $y_t = \mu + \varepsilon_t$, with $\varepsilon_t = \sigma_t z_t$, and $z_t \sim D(0, 1)$ for some distribution D with zero mean and variance one, applying our joint GaMM solution to the moment conditions

$$\mathbf{g}_t = \begin{bmatrix} y_t - \mu \\ (y_t - \mu)^2 - \sigma_t^2 \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ \varepsilon_t^2 - \sigma_t^2 \end{bmatrix}, \quad \mathbf{G}_t = -\mathbf{I}, \quad (22)$$

eq. (20) with $\mathbf{f}_t = \sigma_t^2$ and $\mathbf{C}^B = 0$ reduces to

$$\begin{aligned} \sigma_{t+1}^2 &= \boldsymbol{\omega} + \mathbf{B}(\sigma_t^2 - \boldsymbol{\omega}) + \mathbf{A}(\varepsilon_t^2 - \sigma_t^2) + \mathbf{C}^A \varepsilon_t \\ &= \underbrace{(1 - \mathbf{B}) \boldsymbol{\omega} + A \varepsilon_t^2 + (\mathbf{B} - \mathbf{A}) \sigma_t^2}_{\text{GARCH}(1,1)} + \underbrace{\mathbf{C}^A \varepsilon_t}_{\text{Leverage}}. \end{aligned} \quad (23)$$

For $\mathbf{C}^A = 0$, we obtain the standard updating equation for a generalized autoregressive conditional heteroskedasticity (GARCH) model. For $\mathbf{C}^A \neq 0$, however the additional term $\mathbf{C}^A \varepsilon_t$ captures a leverage type effect: negative returns have a different impact on future volatilities than positive returns. Empirically, returns are often left-skewed which would also translate into an asymmetric response of the volatility to similarly sized negative and positive returns. The effect mimics the leverage effect as specified in the asymmetric AGARCH model of [Engle \(1990\)](#) and [Engle and Ng \(1993\)](#), the optimal filtering specification of [Nelson and Foster \(1994\)](#), and the quadratic QGARCH model of [Sentana \(1995\)](#).

4. An asset-pricing model example with time-varying risk aversion

4.1. The economic model specification

For the development of a non-linear asset pricing model with a time-varying risk aversion coefficient, we adopt the GaMM framework as developed in the earlier sections. Inspired by the seminal work of [Hansen and Singleton \(1982\)](#), we consider a representative agent with a constant relative risk aversion (CRRA) power utility function who faces a consumption-investment problem. This produces the following Euler equation for pricing assets,

$$E_t \left[\beta (C_{t+1}/C_t)^{-\gamma} \mathbf{R}_{t+1}^x \right] = 1, \quad (24)$$

where \mathbf{R}_{t+1}^x denotes the vector of gross asset returns, β is a subjective discount factor, and γ represents the curvature of the utility function. In the case of CRRA utility function, γ is also the relative risk aversion. Both consumption and asset returns are assumed to be expressed in real terms. The Euler equation in eq. (24) implies a stochastic discount factor $M_t = \beta (C_{t+1}/C_t)^{-\gamma}$.

Previous research shows that empirical estimates of the risk aversion parameter γ are sensitive to the particular sample period, starting values, and instruments employed. Furthermore, results of many studies suggest that time series estimates of γ are typically too high compared to risk aversion estimates obtained from experimental data and that the simple model in eq. (24) fails at explaining the cross-sectional variation in stock returns; see, for example, [Savov \(2011\)](#); [Mehra and Prescott \(1985\)](#); [Chen and Ludvigson \(2009\)](#); [Lettau and Ludvigson \(2009\)](#), or [Ludvigson \(2011\)](#) for a summary of the literature and further developments in the field. [Mehra and Prescott \(1985\)](#) referred to this phenomenon as the equity premium puzzle. The reason for poor performance is often attributed to the fact that consumption growth is too smooth relative to the variation in returns and hence the stochastic discount factor needs to be increased by means of high values for the coefficients γ and β .

Poor performance of the standard model in eq. (24) can be addressed in various ways. [Constantinides \(1990\)](#); [Campbell and Cochrane \(1999\)](#), or [Ludvigson \(2011\)](#) discuss allowing for habit formation in the utility specification, which adds an additional source of variation to the stochastic discount factor. Habit models can often be rewritten into eq. (24) with additional multiplicative term. Another strand of literature tries to improve the quality of consumption data that is required

to estimate the model; see, for example, [Savov \(2011\)](#) and [Kroencke \(2017\)](#). These authors argue that consumption data provided by government agencies responsible for compiling macroeconomic statistics² is artificially smoothed. Their main finding is that once smoothing of consumption data is reduced, the static relative risk aversion coefficient is estimated lower. Nonetheless, it is still too high to ‘solve’ the equity premium puzzle in [Mehra and Prescott \(1985\)](#).

At the same time, literature provides strong arguments for allowing the risk aversion parameter to change in time. Stability of deep parameters in the simple structural model in eq. (24) was already questioned by [Ghysels and Hall \(1990\)](#) who introduce a structural break test for γ . We note that [Ghysels and Hall](#) do not find sufficient evidence to reject the hypothesis of a constant risk aversion parameter in their sample, but their tests may have low power against specific mean-reverting alternatives. Furthermore, the aforementioned habit models include time-varying relative risk aversion, but tie it to the dynamics of consumption.³ More recently, [Malmendier and Nagel \(2011\)](#); [Cohn et al. \(2015\)](#); [Guiso et al. \(2018\)](#) provide strong empirical and experimental evidence for changing risk aversion among households, finance professionals, and investors respectively. Collectively these papers attribute these changes to fear stemming from past experiences of agents. For instance, in [Malmendier and Nagel \(2011\)](#) individuals who lived through market busts with low stock returns are less likely to participate in the stock market and are willing to invest a relatively lower proportion of their wealth in risky assets.

Given the evidence for time-varying risk aversion it is interesting to see if the poor performance of the standard model is only a result of mis-specification. As such, we revisit the simple consumption CAPM model in eq. (24), but we allow the relative risk aversion coefficient γ_t to vary over time. We filter out γ_t from the data by endowing the standard Euler equation (24) with GaMM(1,1) dynamics. We assume that the shocks to risk aversion are exogenous and that agents are myopic in the sense that they consider γ_t to remain fixed forever when making their decision at time t . This results in the revised Euler equation

$$E_t \left[\beta (C_{t+1}/C_t)^{-\gamma_t} R_{t+1}^x \right] = 1. \quad (25)$$

Using eq. (25), we can directly employ the GaMM framework by taking $f_t = \gamma_t$.

To estimate \bar{G}_t , which in this case depends on the consumption growth data, the return data, and

²In U.S., the Bureau of Economic Analysis.

³When allowing for habit formation, it is implied that relative risk aversion depends on γ and on the current values of consumption and consumption habit.

on the time-varying parameter \mathbf{f}_t , we use exponentially weighted moving averages. Computing $\bar{\mathbf{G}}_t$ analytically is impossible in this case as we do not know nor do we want to assume the distribution of the risky asset returns and the distribution of consumption growth. Replacing the expectation by a sample average is not appropriate either as the conditional distribution of asset returns given consumption growth in equilibrium that is needed for $\bar{\mathbf{G}}_t$ depends on the current (myopic) risk aversion parameter γ_t . If, however, \mathbf{f}_t varies sufficiently slowly, observations in the recent past can be seen as informative about the curvature $\bar{\mathbf{G}}_t$ of the moment conditions now. Our exponentially weighted moving average (EWMA) estimate of $\bar{\mathbf{G}}_t$ takes the form

$$\hat{\bar{\mathbf{G}}}_t = \lambda \hat{\bar{\mathbf{G}}}_{t-1} + (1 - \lambda) \frac{\partial \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \hat{\boldsymbol{\theta}})}{\partial \mathbf{f}_t^\top}, \quad (26)$$

where we choose λ in the range $(0.98, 1.0)$ and where we initialize the recursion in eq. (26) by $\hat{\bar{\mathbf{G}}}_0 = \mathbf{I}$.⁴ A similar scheme for smoothing the information on the curvature of the score was introduced in [Creal et al. \(2013\)](#).

4.2. Simulation results

For our simulations, we fix $\beta = 1$, which is close to its typical empirical estimate; see for instance [Hansen et al. \(2008\)](#) or [Savov \(2011\)](#). Furthermore, we endow the true risk aversion parameter $\gamma_{0,t}$ either with a structural break or with exogenous AR(1) dynamics. Given a series of $\gamma_{0,t}$, $t = 1, \dots, T$, we use the following data generating process:

$$\Delta c_t = \mu_c + \varepsilon_{ct}, \quad \varepsilon_{ct} \sim N(0, \sigma_\gamma^2), \quad (27)$$

$$R_t = e^{\gamma_{0,t} \Delta c_t} + \varepsilon_{Rt}, \quad \varepsilon_{Rt} \sim N(0, \sigma_R^2), \quad (28)$$

where we set $\mu_c = 0.041$, $\sigma_{ct} = 0.09$, and $\sigma_{Rt} = 0.1$. For the remainder of this paper, we also set $\Omega_{t-1} = \mathbf{I}$. This produces good results in the simulations and the empirical application later on and is also computationally convenient. Estimation results using the GaMM(1,1) specification are presented in Figure 1.

The classical full sample GMM estimates of γ for the structural break and the AR(1) case are

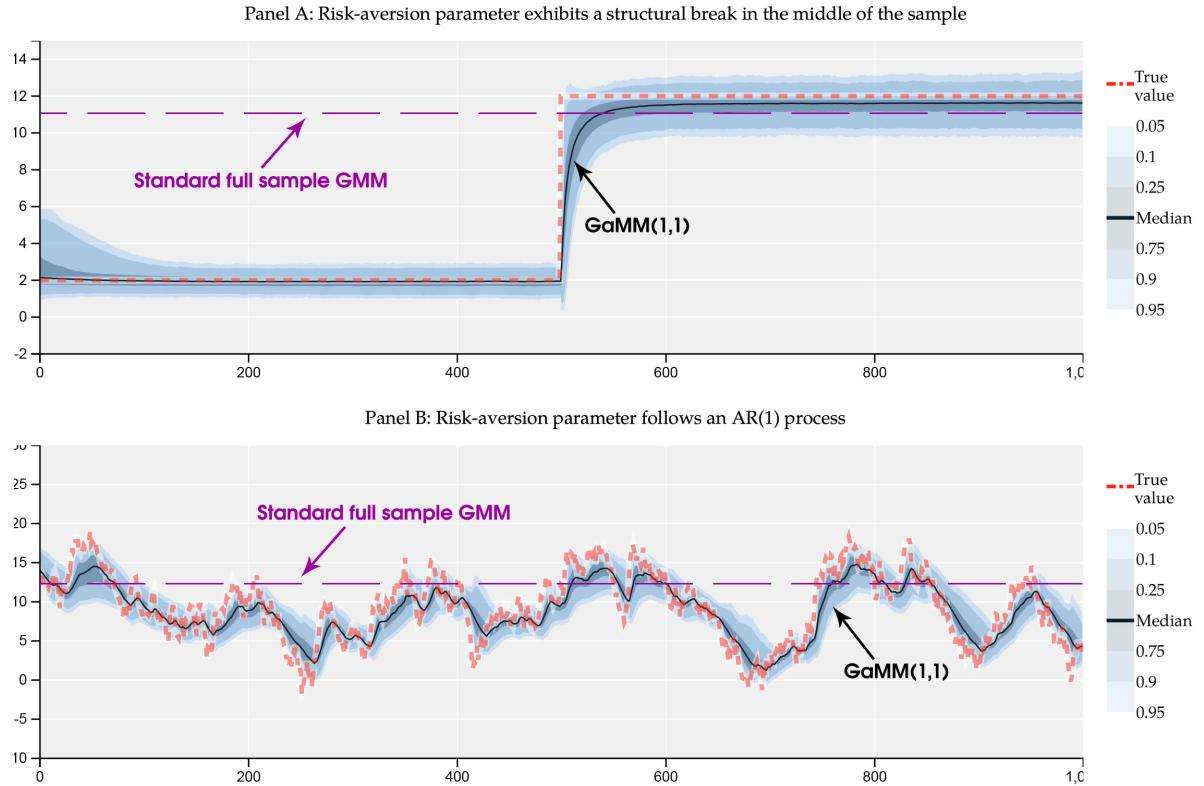
⁴The initial guess $\hat{\bar{\mathbf{G}}}_0 = \mathbf{I}$ can be updated by running the EWMA in reverse, i.e., $\check{\bar{\mathbf{G}}}_t = \lambda \check{\bar{\mathbf{G}}}_{t+1} + (1 - \lambda) \partial \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}) / \partial \mathbf{f}_t$, using $\check{\bar{\mathbf{G}}}_T = \hat{\bar{\mathbf{G}}}_T$ from the forward EWMA as a starting value. Subsequently, we can then re-initialize the forward EWMA by setting $\hat{\bar{\mathbf{G}}}_0 = \check{\bar{\mathbf{G}}}_0$ in eq. (26). This does not increase the computational burden significantly and slightly improves results.

Time-Varying Risk Aversion in CCAPM With Power Utility

This figure illustrates performance of GaMM in estimating the risk aversion parameter in CCAPM. We use the basic power utility specification:

$$1 = E_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma_t} R_{t+1} \right],$$

and assume that agents are myopic about the changes in risk aversion parameter γ_t . Furthermore, in simulations we fix $\beta = 1$ and we do not estimate the discount factor. Results are based on 10,000 replications. Paths were estimated with GaMM(1,1). We juxtapose the true value of the parameter at time t with the median estimate across all replications. We also present 95, 90, and 50% empirical confidence bands constructed based on all simulation results.



11.10 and 12.25, respectively. For the case of the AR(1), this implies that 70% of the γ_t realisations actually lie below the full sample GMM estimate. If there is time-variation in γ_t , the full sample GMM estimates are thus severely biased towards the high-end realizations of the true time-varying risk aversion parameter $\gamma_{0,t}$. As we show later, this may also be part of the explanation of the equity premium puzzle.

If we consider the GaMM(1,1) estimation results, we clearly see that the filtered path $\hat{\gamma}_t$ tends to follow the true path closely. In case of a large structural break, the estimator takes some time to adjust to the new setting. Overall, however, the path is able to capture both the episodes of high and low relative risk aversion. This is a typical result in time-varying parameter models that do not explicitly model a regime switching scenario. In case of the mean-reverting AR(1) dynamics, the GaMM(1,1) approach also recovers the major up and down swings in γ_t .

4.3. Empirical equity premium results for U.S. data

We estimate the model for quarterly U.S. data between 1947 and 2015. As test assets, we use the 3-Month Treasury Bill rate (from the Board of Governors of the Federal Reserve System) and six equity portfolios (double sorted on size and book-to-market as provided by Fama and French). Data on population size in the U.S. as well as expenditure on non-durable goods and consumption of services is provided by the U.S. Bureau of Economic Analysis. Apart from the six Fama-French portfolios we obtain all remaining data from FRED.⁵ All series are deflated with an implicit price deflator (2009=100) which we calculate for the combined consumption (non-durables and services) series. We use two-step feasible GMM/GaMM and report results from the second step only. Standard errors for parameter θ are estimated using the [Newey and West \(1987\)](#) HAC covariance matrix with truncation lag chosen following the procedure proposed by [Newey and West \(1994\)](#). We denote them as $se_{\theta}^{(p)}$. In the results reported below, the model is estimated without additional conditioning information; see [Ludvigson \(2011\)](#) for a discussion of why this is appropriate. We note that adding other instrumental variables to the model does not qualitatively impact the results.

We first estimate the model without time-varying relative risk aversion and obtain a discount rate of $\hat{\beta} = 1.42$ with $se_{\beta}^{(7)} = 0.14$, and a risk aversion parameter of $\hat{\gamma} = 133.06$ with $se_{\gamma}^{(7)} = 38.92$. The high and imprecise estimate for the relative risk aversion is in line with previous research ([Savov, 2011](#); [Lettau and Ludvigson, 2009](#)). A similarly high and unrealistic value of β is reported by [Lettau and Ludvigson \(2009\)](#) for a slightly shorter sample. These results clearly confirm the equity premium puzzle, but also suggest that agents value future utility more than the present one. This is economically unintuitive. A closer inspection of the data suggests that given the static model there are many ‘outlying observations’ (1949–1953, 1960, 1980, and 2008). These observations are

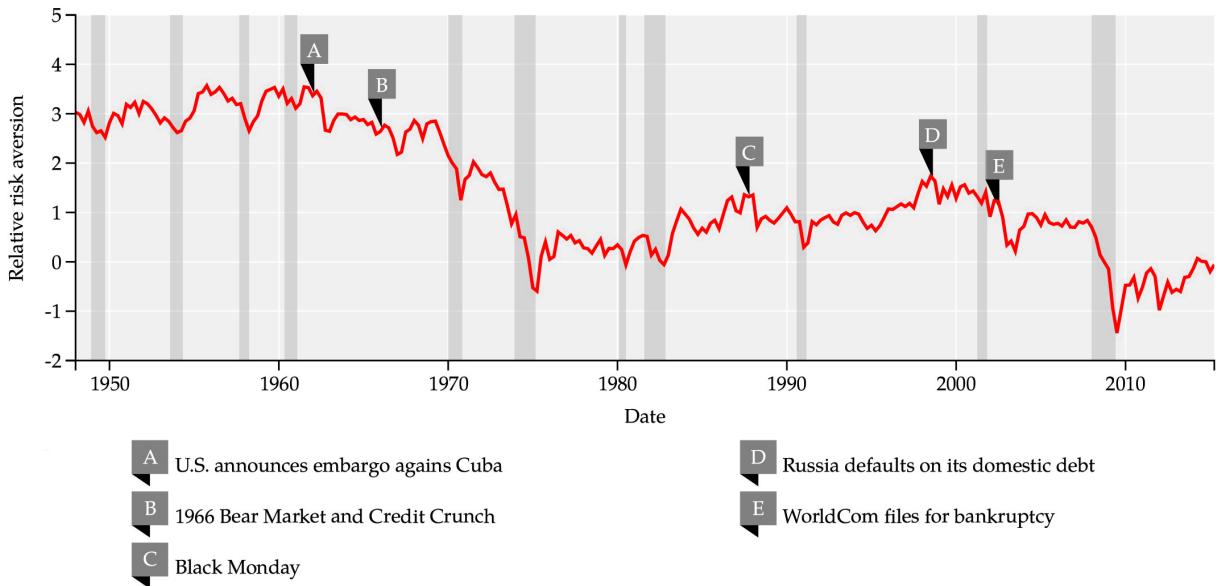
⁵The specific series we use are: 3-Month Treasury Bill (TB3MS), population (B230RC0Q173SBEA), personal consumption expenditures on nondurable goods (PCND) and services (PCESV) with their corresponding implicit price deflators (DNDGRD3Q086SBEA and DSERRD3Q086SBEA respectively).

Equity Premium Puzzle

This figure illustrates performance of GaMM in estimating the time-varying relative risk aversion parameter in CCAPM. We use the basic power utility specification:

$$1 = E_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma_t} R_{t+1} \right],$$

and assume that agents are myopic about the changes in risk aversion parameter γ_t . Tests assets are comprised of the 3-Month Treasury Bill rate and six equity portfolios double sorted on size and book-to-market. We use consumption of non-durable goods and services. All series are deflated with an implicit price deflator (2009=100). Shaded regions correspond to NBER recessions while other relevant events are labelled separately.



both clustered in time and most likely heavily bias the estimates upwards.⁶

In contrast to the static estimation results, GaMM(1,1) produces reasonable values for both the subjective discount factor and the relative risk aversion. The discount factor is estimated at $\hat{\beta} = 0.98$ with $se_{\beta}^{(4)} = 0.001$. Figure 2 shows the estimated path of the relative risk aversion parameter $\hat{\gamma}_t$ based on GaMM(1,1) dynamics together with NBER recession periods. The average value of the

⁶It is worth noting what impact β and γ have on the moment conditions in a neighborhood of time t . *Ceteris paribus*, an increase in the discount rate β increases both the mean and dispersion of the moment conditions without affecting higher moments of their distribution. On the other hand, an increase in the curvature γ increases the variance and skewness of the moment conditions and thus lowers the mean. At the estimated values of the parameters, the ‘outliers’ we identify have an order of magnitude higher contribution to the criterion function and thus have enormous leverage on the outcome.

relative risk aversion coefficient is $\bar{\gamma}_t = 1.42$ (3.13 in 1950–1960, 1.06 in 1990-2000, and 0.55 in 2000-2010). Static parameters governing the GaMM dynamics are estimated at $\hat{\omega} = 3.03$ with $se_{\hat{\omega}}^{(4)} = 0.75$, and $\hat{A} = 0.02$ with $se_{\hat{A}}^{(4)} = 0.005$. The value of \hat{B} is estimated at its boundary value $\hat{B} = 1.00$, in line with the non-stationary downward secular trend in risk aversion $\hat{\gamma}_t$.

Apart from the large downward secular trend in risk aversion, we also see some shorter-term movements. These appear to be cyclical to some extent, though not explicitly modeled as such. Enlarging the model with a cyclical component could be done in the same way as for structural time series models; see [Harvey \(1989\)](#). The short-term cycle appears to follow the business cycle. During recessions and sometimes even before the recession, risk aversion is pushed downwards. Given the postulated utility framework and corresponding Euler equation, these pro-cyclical short-term fluctuations imply that agents adjust their consumption slowly and with a delay compared to reactions of financial markets. In other words, when a recession hits there is a period during which consumption is too high given the observable (negative) returns. In the current limited framework, this can only be explained by a lower risk aversion parameter, which is why we see the drops in γ_t . After the recession ends, we observe risk aversion returning to its long-term trend value. This can happen either because agents adjusted their consumption expenditure or because markets recovered. These patterns are consistent with many of the phenomena and extensions to the basic model as reported in the literature, such as habit formation ([Campbell and Cochrane, 2000](#)), increasing leverage and shortening of investment time-horizon for households ([Adrian and Shin, 2010](#)), or loss aversion ([Kahneman and Tversky, 1984](#)) where agents exhibit (in this case relative) risk-loving behaviour in the loss domain and risk aversion in the gain domain. In line with observations in [Cohn et al. \(2015\)](#) and [Guiso et al. \(2018\)](#) we also observe relative increases of risk aversion after market crashes compared to its value during the bad episodes.

More interesting than the short-term cyclical behaviour is the secular (long-term) pattern in risk aversion. In particular, we notice a continuous decrease in risk aversion since the 1950s, when $\hat{\gamma}_t$ was higher possibly due to the post World War II recovery period which is also consistent with the evidence in [Malmendier and Nagel \(2011\)](#); [Cohn et al. \(2015\)](#); [Guiso et al. \(2018\)](#). Comparing the first and the last decade in the sample shows the extent of the change. On the one hand, the average real consumption growth was between 0.4% and 0.5% in both periods. On the other hand, the real returns in the 1950s were more than three times as high as they were in the 2000s (7% versus 2%) based on data in [Shiller \(2005\)](#). Such a difference in realized returns given the relatively constant pattern of consumption growth can only be consistently explained in the proposed model if risk

aversion has decreased substantially: from the 1950s to the 2000s the GaMM(1,1) estimates of risk aversion decrease from an average value of about 3.13 to a value of about 0.55, i.e., a full order of magnitude. Following the 2008 financial crisis, the model even estimates the representative agent to be risk seeking for some time, roughly up to the end of the European sovereign debt crisis, and risk neutral thereafter. Part of this effect may be due to the return seeking behaviour of institutional investors worldwide in the period of extremely low short term interest rates and non-standard central bank policies in the post-crisis era. Finally, these results do not change qualitatively if we replace the traditional NIPA consumption with data from, e.g., [Kroencke \(2017\)](#).

In short, we conclude that the estimated time-variation in γ_t obtained with the GaMM approach helps us to pinpoint different episodes in U.S. investment behaviour. The estimates obtained are in reasonable ballpark ranges and, in any case, are far from the economically absurd levels obtained for the same utility specification but with static parameters.

5. Conclusions

We have proposed a new approach for modeling time-varying parameters in linear and non-linear econometric models identified through moment conditions. The new approach is referred to as the Generalized autoregressive Method of Moments since it endows parameters that are identified via standard GMM (conditional) moment conditions with autoregressive dynamics based on local deviations of the same (conditional) moment conditions. The method goes substantially beyond previous observation-driven approaches and encompasses many of the previous observation-driven models found in the literature, including the generalized autoregressive score approach of [Creal et al. \(2013\)](#) and [Harvey \(2013\)](#). Given that our method is observation-driven, it also falls directly within the generic GMM framework of [Hansen \(1982\)](#) in terms of the development of the appropriate asymptotic theory for the estimator.

The key novelty in our approach is that we update the time-varying parameters using the influence function for a local (conditional) version of the estimation criterion. Although our current treatment is entirely set in the context of GMM estimation, extensions to other estimation criteria are easily conceived.

We have illustrated our approach in an instrumental variables setting with time-varying regression parameters, and for an asset pricing application using Euler equations. These examples demonstrate that GaMM does not only lead to novel dynamic model specifications, it can also give rise to new

empirical findings. In particular, the secular trending pattern in risk aversion obtained for the standard power utility representative agent model shows that allowing for time variation in deep parameters might be a useful direction to solve some of the empirical puzzles in economics and finance. GaMM dynamics can be viewed as a straightforward and computationally easy tool to uncover potential time-variation in parameters and may therefore provide empirically guidance towards further theoretical model development.

Acknowledgments

We thank Francisco Blasques, Eric Ghysels, Paolo Gorgi, Lars Hansen, Andrew Harvey, Frank Kleibergen, Marc Nientker, Mark Watson. We also thank participants of the Dynamic Models driven by the Score of Predictive Likelihoods Workshops in Amsterdam (Jan 2013) and La Laguna (Jan 2014), the SoFiE Workshop in Cambridge (Apr 2014), the NBER-NSF Time Series Conference at the FED St. Louis (Sep 2014), the Econometric Society European Winter Meeting (Dec 2014), the 8th Annual SoFiE Conference in Aarhus (Jun 2015), the Montréal 2015 World Congress of the Econometric Society (Aug 2015), the 2015 European Finance Association Annual Meeting in Vienna (Aug 2015), the European Economic Society 2015 meeting in Mannheim (Aug 2015), the EC² meeting in Amsterdam (Dec 2017), the Rimini Time Series Conference (Jun 2018), and seminar participants at Warwick Business School (2017) for helpful comments. All numerical results presented in this paper were obtained with OxMetrics 7/Python/Julia at the Dutch National Cluster (LISA).

References

- Adrian, T. and H. S. Shin (2010). Liquidity and leverage. *Journal of Financial Intermediation* 19(3), 418–437.
- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* 61(4), 821–856.
- Blasques, F., S. J. Koopman, and A. Lucas (2014). Maximum likelihood estimation for Generalized Autoregressive Score models. *Tinbergen Institute Discussion Paper* (TI 14-029/III).
- Blasques, F., S. J. Koopman, and A. Lucas (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika* 102(2), 325–343.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Campbell, J. Y. and J. Cochrane (1999, April). Force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107(2), 205–251.

Campbell, J. Y. and J. H. Cochrane (2000). Explaining the poor performance of consumption-based asset pricing models. *The Journal of Finance* 55(6), 2863–2878.

Chen, X. and S. C. Ludvigson (2009). Land of addicts? an empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics* 24(7), 1057–1093.

Cohn, A., J. Engelmann, E. Fehr, and M. A. Maréchal (2015, February). Evidence for countercyclical risk aversion: An experiment with financial professionals. *American Economic Review* 105(2), 860–85.

Constantinides, G. M. (1990). Habit formation: A resolution of the equity premium puzzle. *Journal of political Economy*, 519–543.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8, 93–115.

Cox, D. R. and D. V. Hinkley (1979). *Theoretical statistics*. CRC Press.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics* 29(4).

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28(5), 777–795.

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation driven mixed-measurement dynamic factor models with an application to credit risk. *Review of Economics and Statistics* 96(5), 898–915.

Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*. Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.

Engle, R. F. (1990). Stock volatility and the crash of 87: Discussion. *The Review of Financial Studies* 3(1), 103–106.

Engle, R. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance* 48(5), 1749–1778.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66(5), 1127–1162.

Fernández-Villaverde, J. and J. Rubio-Ramírez (2013). Macroeconomics and Volatility: Data, Models, and Methods. In D. Acemoglu, M. Arellano, and E. Dekel (Eds.), *Advances in Economics and Econometrics: Tenth World Congress*, pp. 137–183. Cambridge, UK: Cambridge University Press.

Ghysels, E. and A. Hall (1990). Are consumption-based intertemporal capital asset pricing models structural? *Journal of Econometrics* 45(1), 121–139.

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

Guiso, L., P. Sapienza, and L. Zingales (2018). Time varying risk aversion. *Journal of Financial Economics* 128(3), 403–421.

Hamilton, J. D. (2010). Macroeconomics and ARCH. In T. Bollerslev, J. R. Russell, and M. Watson (Eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, pp. 79–95. Cambridge, UK: Cambridge University Press.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (2011). *Robust statistics: the approach based on influence functions*, Volume 114. John Wiley & Sons.

Hansen, B. E. and K. D. West (2002). Generalized Method of Moments and macroeconomics. *Journal of Business and Economic Statistics* 20(4), 460–469.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–54.

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

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

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge, UK: Cambridge University Press.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails*. Cambridge University Press.

Harvey, A. C. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association* 507, 1112–1122.

Kahneman, D. and A. Tversky (1984). Choices, values, and frames. *American Psychologist* 39(4), 341.

Kroencke, T. A. (2017). Asset pricing without garbage. *The Journal of Finance* 72(1), 47–98.

Kuersteiner, G. M. (2012). Kernel-weighted GMM estimators for linear time series models. *Journal of Econometrics* 170(2), 399–421.

Lettau, M. and S. C. Ludvigson (2009). Euler equation errors. *Review of Economic Dynamics* 12(2), 255–283.

Lucas, A., J. Schaumburg, and B. Schwaab (2018). Bank business models at zero interest rates. *Journal of Business and Economic Statistics*, in press.

Ludvigson, S. C. (2011). Advances in consumption-based asset pricing: Empirical tests. Technical report, National Bureau of Economic Research.

Malmendier, U. and S. Nagel (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics* 126(1), 373–416.

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

Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics* 45(1-2), 7–38.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 347–370.

- Nelson, D. B. and D. P. Foster (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica* 62(1), 1–41.
- Newey, W. K. and K. D. West (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777–787.
- Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies* 61(4), 631–653.
- Oh, D. H. and A. Patton (2018). Time-varying systemic risk: evidence from a dynamic copula model of cds spreads. *Journal of Business and Economic Statistics*, in press.
- Opschoor, A., P. Janus, A. Lucas, and D. J. van Dijk (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business and Economic Statistics*, in press.
- Patton, A. J., J. F. Ziegel, and R. Chen (2017). Dynamic semiparametric models for Expected Shortfall (and Value-at-Risk). *Arxiv preprint https://arxiv.org/abs/1707.05108*.
- Ronchetti, E. and F. Trojani (2001). Robust inference with GMM estimators. *Journal of econometrics* 101(1), 37–69.
- Savov, A. (2011). Asset pricing with garbage. *The Journal of Finance* 66(1), 177–201.
- Sentana, E. (1995). Quadratic ARCH models. *The Review of Economic Studies* 62(4), 639–661.
- Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford, UK: Oxford University Press.
- Shiller, R. (2005). *Irrational Exuberance*. Broadway Books. Currency/Doubleday.
- Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroeskedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics* 34(5), 2449–2495.
- White, H. (1996). *Estimation, inference and specification analysis*. Number 22. Cambridge university press.
- Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH(1, 1) model. *Scandinavian Journal of Statistics* 40(4), 846–867.

Supplementary Appendix to: Generalized autoregressive Method of Moments

Drew Creal, Siem Jan Koopman, André Lucas, Marcin Zamojski

Contents

Appendix A: Proofs	S.2
Appendix B: Time-varying linear regression models with endogenous covariates	S.7

Appendix A: Proofs

Proof of Proposition 1. The result can be found directly in for instance Ronchetti and Trojani (2001). For completeness, we provide the standard derivation following Hampel et al. (2011). Taking derivatives of eq. (7) with respect of ϵ , we have

$$\begin{aligned}
0 &= \frac{d}{d\epsilon} \left(\bar{\mathbf{G}}_{f,t}^\epsilon \right)^\top \Omega_{t-1} E_{t-1}^\epsilon [\mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)] \Big|_{\epsilon=0} \\
&= \left(\bar{\mathbf{G}}_{f,t}^\epsilon \right)^\top \Omega_{t-1} \cdot \frac{d}{d\epsilon} E_{t-1}^\epsilon [\mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)] \Big|_{\epsilon=0} \\
&= \left(\bar{\mathbf{G}}_{f,t}^\epsilon \right)^\top \Omega_{t-1} \cdot \frac{d}{d\epsilon} \left(\int \mathbf{g}_t(\mathbf{w}; \mathbf{f}_t(F_w^\epsilon), \boldsymbol{\theta}_1) d((1-\epsilon)F_w + \epsilon\Delta_{w_t}) \right) \Big|_{\epsilon=0} \\
&= \left(\bar{\mathbf{G}}_{f,t}^\epsilon \right)^\top \Omega_{t-1} \cdot \left(\int \mathbf{g}_t(\mathbf{w}; \mathbf{f}_t(F_w), \boldsymbol{\theta}_1) d(\Delta_{w_t} - F_w) + \bar{\mathbf{G}}_{f,t}^0 \frac{d\mathbf{f}_t(F_w^\epsilon)}{d\epsilon} \Big|_{\epsilon=0} \right) \\
&= \left(\bar{\mathbf{G}}_{f,t}^\epsilon \right)^\top \Omega_{t-1} \cdot \left(\mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t(F_w), \boldsymbol{\theta}_1) + \bar{\mathbf{G}}_{f,t}^0 \frac{d\mathbf{f}_t(F_w^\epsilon)}{d\epsilon} \Big|_{\epsilon=0} \right) \Leftrightarrow \\
\frac{d\mathbf{f}_t(F_w^\epsilon)}{d\epsilon} \Big|_{\epsilon=0} &= - \left(\left(\bar{\mathbf{G}}_{f,t}^\epsilon \right)^\top \Omega_{t-1} \bar{\mathbf{G}}_{f,t} \right)^{-1} \left(\bar{\mathbf{G}}_{f,t}^\epsilon \right)^\top \Omega_{t-1} \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1),
\end{aligned}$$

with $\bar{\mathbf{G}}_{f,t} = \bar{\mathbf{G}}_{f,t}^0$, and where the second equality follows from eq. (5); see also Hampel et al. (2011) for more details. \square

Proof of Proposition 2. Define the time t conditional expectations

$$\bar{\mathbf{g}}_t(\mathbf{f}) = \int \mathbf{g}_t(\mathbf{w}; \mathbf{f}, \boldsymbol{\theta}_1) dF_w(\mathbf{w}), \quad (\text{A.1})$$

$$\bar{\mathbf{G}}_{f,t}(\mathbf{f}) = \int \frac{\partial \mathbf{g}_t(\mathbf{w}; \mathbf{f}, \boldsymbol{\theta}_1)}{\partial \mathbf{f}^\top} dF_w(\mathbf{w}), \quad (\text{A.2})$$

where $F_w(w)$ is the *unknown* true conditional distribution of \mathbf{w}_t given all information up to time $t-1$. Define $\bar{\mathbf{V}}_t = \bar{\mathbf{G}}_{f,t}(\mathbf{f}_t)^\top \Omega_{t-1} \bar{\mathbf{G}}_{f,t}(\mathbf{f}_t)$ and a GaMM(1,1) step with $\omega = 0$, $\mathbf{A} = \mathbf{a}\mathbf{I}$ for $\mathbf{a} \in \mathbb{R}^+$, and $\mathbf{B} = \mathbf{I}$. Under the current differentiability conditions, we obtain

$$\begin{aligned}
\bar{\mathbf{g}}_t(\mathbf{f}') &= \bar{\mathbf{g}}_t(\mathbf{f}) + \bar{\mathbf{G}}_{f,t}(\mathbf{f}^*) (\mathbf{f}' - \mathbf{f}), \\
\bar{\mathbf{G}}_{f,t}(\mathbf{f}') &= \bar{\mathbf{G}}_{f,t}(\mathbf{f}) + \sum_{j=1}^{\dim(\mathbf{f})} \frac{\partial \bar{\mathbf{G}}_{f,t}(\mathbf{f})}{\partial \mathbf{f}_j} \Big|_{\mathbf{f}=\mathbf{f}^{**}} \cdot (\mathbf{f}' - \mathbf{f})_j,
\end{aligned}$$

where \mathbf{f}^* and \mathbf{f}^{**} are points on the line segment between \mathbf{f} and \mathbf{f}' , and $\dim(\mathbf{f})$ denotes the dimension of \mathbf{f} . Using these two expansions, and defining $\Delta\mathbf{f}_{t+1} = \mathbf{f}_{t+1} - \mathbf{f}_t$ and $\partial_j \bar{\mathbf{G}}_{f,t}(\mathbf{f}) =$

$\partial \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f})/\partial \mathbf{f}_j$, we have

$$\begin{aligned}
& \mathcal{C}(t, \mathbf{f}_t) - \mathcal{C}(t, \mathbf{f}_{t+1}) \\
&= \bar{\mathbf{g}}_t(\mathbf{f}_t)^\top \Omega_{t-1} \bar{\mathbf{g}}_t(\mathbf{f}_t) - \bar{\mathbf{g}}_t(\mathbf{f}_{t+1})^\top \Omega_{t-1} \bar{\mathbf{g}}_t(\mathbf{f}_{t+1}) \\
&= \bar{\mathbf{g}}_t(\mathbf{f}_t)^\top \Omega_{t-1} \bar{\mathbf{g}}_t(\mathbf{f}_t) - \\
&\quad (\bar{\mathbf{g}}_t(\mathbf{f}_t) + \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^*) \Delta \mathbf{f}_{t+1})^\top \Omega_{t-1} (\bar{\mathbf{g}}_t(\mathbf{f}_t) + \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^*) \Delta \mathbf{f}_{t+1}) \\
&= -2\Delta \mathbf{f}_{t+1}^\top \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^*)^\top \Omega_{t-1} \bar{\mathbf{g}}_t(\mathbf{f}_t) \\
&\quad - \Delta \mathbf{f}_{t+1}^\top \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^*)^\top \Omega_{t-1} \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^*) \Delta \mathbf{f}_{t+1} \\
&= -2\Delta \mathbf{f}_{t+1}^\top \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_t)^\top \Omega_{t-1} \bar{\mathbf{g}}_t(\mathbf{f}_t) \\
&\quad - 2\Delta \mathbf{f}_{t+1}^\top \left(\sum_{j=1}^{\dim(\mathbf{f})} \partial_j \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^{**}) \cdot (\mathbf{f}_{t+1}^* - \mathbf{f}_t)_j \right)^\top \Omega_{t-1} \bar{\mathbf{g}}_t(\mathbf{f}_t) \\
&\quad - \Delta \mathbf{f}_{t+1}^\top \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_t)^\top \Omega_{t-1} \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_t) \Delta \mathbf{f}_{t+1} \\
&\quad - 2\Delta \mathbf{f}_{t+1}^\top \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_t)^\top \Omega_{t-1} \left(\sum_{j=1}^{\dim(\mathbf{f})} \partial_j \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^{**}) \cdot (\mathbf{f}_{t+1}^* - \mathbf{f}_t)_j \right) \Delta \mathbf{f}_{t+1} \\
&\quad - \Delta \mathbf{f}_{t+1}^\top \left(\sum_{j=1}^{\dim(\mathbf{f})} \partial_j \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^{**}) \cdot (\mathbf{f}_{t+1}^* - \mathbf{f}_t)_j \right)^\top \Omega_{t-1} \left(\sum_{j=1}^{\dim(\mathbf{f})} \partial_j \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^{**}) \cdot (\mathbf{f}_{t+1}^* - \mathbf{f}_t)_j \right) \Delta \mathbf{f}_{t+1} \\
&= -2\mathbf{a} \mathbf{g}_t^\top \Omega_{t-1} \bar{\mathbf{G}}_{\mathbf{f},t} \bar{\mathbf{V}}_t \bar{\mathbf{G}}_{\mathbf{f},t}^\top \Omega_{t-1} \bar{\mathbf{g}}_t \\
&\quad - 2\mathbf{a} \mathbf{a}^* \mathbf{g}_t^\top \Omega_{t-1} \bar{\mathbf{G}}_{\mathbf{f},t} \bar{\mathbf{V}}_t \left(\sum_{j=1}^{\dim(\mathbf{f})} \partial_j \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^{**}) \cdot (\bar{\mathbf{V}}_t \bar{\mathbf{G}}_{\mathbf{f},t}^\top \Omega_{t-1} \mathbf{g}_t)_j \right)^\top \Omega_{t-1} \bar{\mathbf{g}}_t \\
&\quad - \mathbf{a}^2 \mathbf{g}_t^\top \Omega_{t-1} \bar{\mathbf{G}}_{\mathbf{f},t} \bar{\mathbf{V}}_t \bar{\mathbf{G}}_{\mathbf{f},t}^\top (\mathbf{f}_t) \Omega_{t-1} \bar{\mathbf{G}}_{\mathbf{f},t} \bar{\mathbf{V}}_t \bar{\mathbf{G}}_{\mathbf{f},t}^\top \Omega_{t-1} \mathbf{g}_t \\
&\quad - 2\mathbf{a}^2 \mathbf{a}^* \mathbf{g}_t^\top \Omega_{t-1} \bar{\mathbf{G}}_{\mathbf{f},t} \bar{\mathbf{V}}_t \bar{\mathbf{G}}_{\mathbf{f},t}^\top \Omega_{t-1} \left(\sum_{j=1}^{\dim(\mathbf{f})} \partial_j \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^{**}) \cdot (\bar{\mathbf{V}}_t \bar{\mathbf{G}}_{\mathbf{f},t}^\top \Omega_{t-1} \mathbf{g}_t)_j \right) \bar{\mathbf{V}}_t \bar{\mathbf{G}}_{\mathbf{f},t}^\top \Omega_{t-1} \mathbf{g}_t \\
&\quad - \mathbf{a}^2 (\mathbf{a}^*)^2 \mathbf{g}_t^\top \Omega_{t-1} \bar{\mathbf{G}}_{\mathbf{f},t} \bar{\mathbf{V}}_t \left(\sum_{j=1}^{\dim(\mathbf{f})} \partial_j \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^{**}) \cdot (\bar{\mathbf{V}}_t \bar{\mathbf{G}}_{\mathbf{f},t}^\top \Omega_{t-1} \mathbf{g}_t)_j \right)^\top \Omega_{t-1} \times \\
&\quad \left(\sum_{j=1}^{\dim(\mathbf{f})} \partial_j \bar{\mathbf{G}}_{\mathbf{f},t}(\mathbf{f}_{t+1}^{**}) \cdot (\bar{\mathbf{V}}_t \bar{\mathbf{G}}_{\mathbf{f},t}^\top \Omega_{t-1} \mathbf{g}_t)_j \right) \bar{\mathbf{V}}_t \bar{\mathbf{G}}_{\mathbf{f},t}^\top \Omega_{t-1} \mathbf{g}_t
\end{aligned}$$

where \mathbf{f}_{t+1}^* and \mathbf{f}_{t+1}^{**} are points between \mathbf{f}_t and \mathbf{f}_{t+1} , \mathbf{a}^* is a scalar between \mathbf{a} and 0, $\mathbf{g}_t = \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \theta_1)$, $\bar{\mathbf{g}}_t = \bar{\mathbf{g}}_t(\mathbf{f}_t)$, and $\bar{\mathbf{G}}_t = \bar{\mathbf{G}}_t(\mathbf{f}_t)$. Taking conditional expectations, and using the existence of fourth order moments and the finiteness of all relevant weight and derivative matrices,

we obtain

$$\mathbb{E}_{t-1} [\mathcal{C}(t, \mathbf{f}_t) - \mathcal{C}(t, \mathbf{f}_{t+1})] = -2\mathbf{a}\bar{\mathbf{g}}_t^\top \boldsymbol{\Omega}_{t-1} \bar{\mathbf{G}}_{f,t} \bar{\mathbf{V}}_t \bar{\mathbf{G}}_{f,t}^\top \boldsymbol{\Omega}_{t-1} \bar{\mathbf{g}}_t + \mathcal{O}(\mathbf{a}^2), \quad (\text{A.3})$$

where $\mathcal{O}(\mathbf{a}^2)$ denotes a term vanishing at order \mathbf{a}^2 . It is then clear that for small enough \mathbf{a} , this expected difference is positive, which proves the result. \square

Proof of Proposition 3. Let $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)$, with $\boldsymbol{\theta}_2^\top = (\boldsymbol{\omega}^\top, \text{vec}[\mathbf{B}]^\top, \text{vec}[\mathbf{A}]^\top)$ containing the static parameters governing the GaMM(1,1) dynamics, and $\boldsymbol{\theta}_1$ containing the remaining static parameters. Taking the total derivative of the moment function, we obtain

$$\frac{d\mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)}{d\boldsymbol{\theta}^\top} = \frac{\partial \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)}{\partial \mathbf{f}_t^\top} \cdot \frac{d\mathbf{f}_t(\mathbf{w}_{t-1}; \mathbf{f}_{t-1}, \boldsymbol{\theta})}{d\boldsymbol{\theta}^\top},$$

such that

$$\mathbf{Z}_t^\top = \boldsymbol{\Omega}_{t-1} \mathbf{G}_{\boldsymbol{\theta}, t} + \boldsymbol{\Omega}_{t-1} \mathbf{G}_{f,t} \frac{d\mathbf{f}_t(\mathbf{w}_{t-1}; \mathbf{f}_{t-1}, \boldsymbol{\theta})}{d\boldsymbol{\theta}^\top}, \quad (\text{A.4})$$

with

$$\mathbf{G}_{\boldsymbol{\theta}, t} = \mathbb{E}_{t-1} \left[\frac{\partial \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}^\top} \right] = \left(\begin{array}{c|c} \mathbb{E}_{t-1} \left[\frac{\partial \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1^\top} \right] & 0 \\ \hline \mathbf{G}_{\boldsymbol{\theta}_1, t} & 0 \end{array} \right),$$

and $\mathbf{G}_t = \mathbf{G}_t^0$ as defined in eq. (7). The last derivative in eq. (A.4) follows the recursion

$$\begin{aligned} \frac{d\mathbf{f}_{t+1}}{d\boldsymbol{\theta}^\top} &= \frac{d\mathbf{f}_{t+1}(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta})}{d\boldsymbol{\theta}^\top} \\ &= \frac{d(\mathbf{I} - \mathbf{B})\boldsymbol{\omega}}{d\boldsymbol{\theta}^\top} + (\mathbf{f}_t(\mathbf{w}_{t-1}; \mathbf{f}_{t-1}, \boldsymbol{\theta})^\top \otimes \mathbf{I}) \frac{d \text{vec}[\mathbf{B}]}{d\boldsymbol{\theta}^\top} + \\ &\quad (\mathbf{s}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta})^\top \otimes \mathbf{I}) \frac{d \text{vec}[\mathbf{A}]}{d\boldsymbol{\theta}^\top} + \mathbf{B} \frac{d\mathbf{f}_t(\mathbf{w}_{t-1}; \mathbf{f}_{t-1}, \boldsymbol{\theta})}{d\boldsymbol{\theta}^\top} + \\ &\quad \mathbf{A} \frac{\partial \mathbf{s}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} + \mathbf{A} \frac{\partial \mathbf{s}_t(\mathbf{w}_t; \mathbf{f}_t, \boldsymbol{\theta})}{\partial \mathbf{f}_t^\top} \frac{d\mathbf{f}_t(\mathbf{w}_{t-1}; \mathbf{f}_{t-1}, \boldsymbol{\theta})}{d\boldsymbol{\theta}^\top} \\ &= \left(\begin{array}{c|c} 0 & \mathbf{I} - \mathbf{B} \\ \hline \mathbf{f}_t^\top \otimes \mathbf{I} & \mathbf{s}_t^\top \otimes \mathbf{I} \end{array} \right) + \mathbf{B} \frac{d\mathbf{f}_t}{d\boldsymbol{\theta}^\top} + \mathbf{A} \frac{\partial \mathbf{s}_t}{\partial \boldsymbol{\theta}^\top} + \mathbf{A} \frac{\partial \mathbf{s}_t}{\partial \mathbf{f}_t^\top} \frac{d\mathbf{f}_t}{d\boldsymbol{\theta}^\top}. \end{aligned} \quad (\text{A.5})$$

In the first term in eq. (A.5), we directly recognize the instruments proposed in equation eq. (15). In empirical work, \mathbf{A} is typically small, while \mathbf{B} is large with roots close to the unit circle. Using the propositions premise that $\mathbf{B} = (1 - \mathbf{a}) \mathbf{I}$ and $\mathbf{A} = \mathbf{a} \mathbf{I}$ for small \mathbf{a} , it follows immediately from

eq. (A.5) that we can write

$$\frac{d\mathbf{f}_{t+1}}{d\boldsymbol{\theta}^\top} = \sum_{i=0}^{t-1} (1-\alpha)^i \cdot \begin{pmatrix} 0 & \mathbf{I} - \mathbf{B} & \mathbf{f}_{t-i}^\top \otimes \mathbf{I} & \mathbf{s}_{t-i}^\top \otimes \mathbf{I} \end{pmatrix} + \mathcal{O}(\alpha),$$

which proves the result. \square

Proof of Proposition 4. To establish consistency, the main difficulty is that the initialized series $\mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$ is not stationary and ergodic. We show however that we can replace $\mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$ by its stationary and ergodic uninitialized limit $\mathbf{f}_t(\boldsymbol{\theta})$, after which we can appeal to the standard argument for the consistency of the GMM estimator for stationary and ergodic data, where the data in this case consists of the sequence $\{(\mathbf{w}_t, \mathbf{f}_t(\boldsymbol{\theta}))\}_{t \in \mathbb{Z}}$.

We write the sample moment condition as

$$\begin{aligned} \tilde{\mathbf{g}}_T(\mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1), \boldsymbol{\theta}) \\ = \frac{1}{T} \sum_{t=1}^T \mathbf{Z}_t(\mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)) \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1), \boldsymbol{\theta}_1) \\ = \frac{1}{T} \sum_{t=1}^T \mathbf{Z}_t(\mathbf{f}_t(\boldsymbol{\theta})) \mathbf{g}_t((\mathbf{w}_t; \mathbf{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_1) + \\ \frac{1}{T} \sum_{t=1}^T (\mathbf{Z}_t(\mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)) \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1), \boldsymbol{\theta}_1) - \mathbf{Z}_t(\mathbf{f}_t(\boldsymbol{\theta})) \mathbf{g}_t(\mathbf{w}_t; \mathbf{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_1)). \end{aligned} \quad (\text{A.6})$$

The first term converges to its expectation using Assumption B.2 and the ergodic theorem. The second line in eq. (A.6) converges to zero almost surely by Assumption B.1, the continuity of \mathbf{Z}_t and \mathbf{g}_t in \mathbf{f}_t , and the continuous mapping theorem. The remainder of the consistency proof now follows the standard line of argument as found in for instance White (1996).

For asymptotic normality, we consider the first order condition

$$0 = \left(\frac{d\tilde{\mathbf{g}}_T(\mathbf{f}_t(\hat{\boldsymbol{\theta}}), \bar{\mathbf{f}}_1), \hat{\boldsymbol{\theta}})}{d\boldsymbol{\theta}^\top} \right)^\top \tilde{\boldsymbol{\Omega}}_T \tilde{\mathbf{g}}_T(\mathbf{f}_t(\hat{\boldsymbol{\theta}}, \bar{\mathbf{f}}_1), \hat{\boldsymbol{\theta}}),$$

and expand it around $\boldsymbol{\theta}_0^*$ as

$$0 = \left(\frac{d\tilde{\mathbf{g}}_T(\mathbf{f}_t(\boldsymbol{\theta}_0^*, \bar{\mathbf{f}}_1), \boldsymbol{\theta}_0^*)}{d\boldsymbol{\theta}^\top} \right)^\top \tilde{\boldsymbol{\Omega}}_T \tilde{\mathbf{g}}_T(\mathbf{f}_t(\boldsymbol{\theta}_0^*, \bar{\mathbf{f}}_1), \boldsymbol{\theta}_0^*) + \quad (\text{A.7})$$

$$\sum_{i=1}^{\dim(\boldsymbol{\theta})} \left[\left(\frac{d^2 \tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1), \bar{\boldsymbol{\theta}}^\star)}{d\boldsymbol{\theta}^\top d\boldsymbol{\theta}_i} \right)^\top \tilde{\boldsymbol{\Omega}}_T \tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1), \bar{\boldsymbol{\theta}}^\star) + \left(\frac{d\tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1), \bar{\boldsymbol{\theta}}^\star)}{d\boldsymbol{\theta}^\top} \right)^\top \tilde{\boldsymbol{\Omega}}_T \frac{d\tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1), \bar{\boldsymbol{\theta}}^\star)}{d\boldsymbol{\theta}_i} \right] \times (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{0,i}^\star), \quad (\text{A.8})$$

where $\bar{\boldsymbol{\theta}}^\star$ is a point between $\boldsymbol{\theta}_0^\star$ and $\hat{\boldsymbol{\theta}}$. From Assumption B.2 we have $\tilde{\boldsymbol{\Omega}}_T \xrightarrow{\text{a.s.}} \tilde{\boldsymbol{\Omega}}$. We also note that

$$\frac{d\tilde{\mathbf{g}}_T (\mathbf{f}_t(\boldsymbol{\theta}_0^\star, \bar{\mathbf{f}}_1), \boldsymbol{\theta}_0^\star)}{d\boldsymbol{\theta}^\top} = \frac{\partial \tilde{\mathbf{g}}_T (\mathbf{f}_t(\boldsymbol{\theta}_0^\star, \bar{\mathbf{f}}_1), \boldsymbol{\theta}_0^\star)}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \tilde{\mathbf{g}}_T (\mathbf{f}_t(\boldsymbol{\theta}_0^\star, \bar{\mathbf{f}}_1), \boldsymbol{\theta}_0^\star)}{\partial \mathbf{f}^\top} \frac{\partial \mathbf{f}_t(\boldsymbol{\theta}_0^\star, \bar{\mathbf{f}}_1)}{\partial \boldsymbol{\theta}^\top} \quad (\text{A.9})$$

$$\xrightarrow{\text{a.s.}} \frac{\partial \tilde{\mathbf{g}}_T (\mathbf{f}_t(\boldsymbol{\theta}_0^\star), \boldsymbol{\theta}_0^\star)}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \tilde{\mathbf{g}}_T (\mathbf{f}_t(\boldsymbol{\theta}_0^\star), \boldsymbol{\theta}_0^\star)}{\partial \mathbf{f}^\top} \frac{\partial \mathbf{f}_t(\boldsymbol{\theta}_0^\star)}{\partial \boldsymbol{\theta}^\top}, \quad (\text{A.10})$$

as $\boldsymbol{\theta}$ affects $\tilde{\mathbf{g}}_T$ both directly and via the time-varying parameter $\mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$. Note that eq. (A.10) follows from eq. (A.9) via the continuous mapping theorem and Assumption B.1, which holds for both $\mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$ and its first two derivatives. Also note that via the continuous mapping theorem and the *exponentially fast* almost sure convergence of $\mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$ to $\mathbf{f}_t(\boldsymbol{\theta})$, we obtain

$$\sqrt{T} \cdot \left(\tilde{\mathbf{g}}_T (\mathbf{f}_t(\boldsymbol{\theta}_0^\star, \bar{\mathbf{f}}_1), \boldsymbol{\theta}_0^\star) - \tilde{\mathbf{g}}_T (\mathbf{f}_t(\boldsymbol{\theta}_0^\star), \boldsymbol{\theta}_0^\star) \right) \xrightarrow{\text{a.s.}} 0, \quad (\text{A.11})$$

such that eq. (A.9) follows a central limiting result from Assumption B.5 along the standard lines of argument; see for instance White (1996).

From eq. (A.9), Assumption B.1, and the continuous mapping theorem, we also obtain

$$\begin{aligned} & \frac{d^2 \tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1), \bar{\boldsymbol{\theta}}^\star)}{d\boldsymbol{\theta}^\top d\boldsymbol{\theta}_i} = \\ & \frac{\partial^2 \tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1), \bar{\boldsymbol{\theta}}^\star)}{\partial \boldsymbol{\theta}^\top \partial \boldsymbol{\theta}_i} + \sum_{j=1}^{\dim(f)} \frac{\partial^2 \tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1), \bar{\boldsymbol{\theta}}^\star)}{\partial \boldsymbol{\theta}^\top \partial \mathbf{f}_j} \frac{\partial \mathbf{f}_{t,j}(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1)}{\partial \boldsymbol{\theta}_i} + \\ & \frac{\partial^2 \tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1), \bar{\boldsymbol{\theta}}^\star)}{\partial \mathbf{f}^\top \partial \boldsymbol{\theta}_i} \frac{\partial \mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1)}{\partial \boldsymbol{\theta}^\top} + \sum_{j=1}^{\dim(f)} \frac{\partial^2 \tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1), \bar{\boldsymbol{\theta}}^\star)}{\partial \mathbf{f}^\top \partial \mathbf{f}_j} \frac{\partial \mathbf{f}_{t,j}(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1)}{\partial \boldsymbol{\theta}_i} \frac{\partial \mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1)}{\partial \boldsymbol{\theta}^\top} + \\ & \frac{\partial \tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1), \bar{\boldsymbol{\theta}}^\star)}{\partial \mathbf{f}^\top} \frac{\partial^2 \mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star, \bar{\mathbf{f}}_1)}{\partial \boldsymbol{\theta}^\top \partial \boldsymbol{\theta}_i} \\ & \xrightarrow{\text{a.s.}} \frac{\partial^2 \tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star), \bar{\boldsymbol{\theta}}^\star)}{\partial \boldsymbol{\theta}^\top \partial \boldsymbol{\theta}_i} + \sum_{j=1}^{\dim(f)} \frac{\partial^2 \tilde{\mathbf{g}}_T (\mathbf{f}_t(\bar{\boldsymbol{\theta}}^\star), \bar{\boldsymbol{\theta}}^\star)}{\partial \boldsymbol{\theta}^\top \partial \mathbf{f}_j} \frac{\partial \mathbf{f}_{t,j}(\bar{\boldsymbol{\theta}}^\star)}{\partial \boldsymbol{\theta}_i} + \end{aligned}$$

$$\begin{aligned} & \frac{\partial^2 \tilde{\mathbf{g}}_T(\mathbf{f}_t(\bar{\boldsymbol{\theta}}^*), \bar{\boldsymbol{\theta}}^*)}{\partial \mathbf{f}^\top \partial \boldsymbol{\theta}_i} \frac{\partial \mathbf{f}_t(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^\top} + \sum_{j=1}^{\dim(f)} \frac{\partial^2 \tilde{\mathbf{g}}_T(\mathbf{f}_t(\bar{\boldsymbol{\theta}}^*), \bar{\boldsymbol{\theta}}^*)}{\partial \mathbf{f}^\top \partial \mathbf{f}_j} \frac{\partial \mathbf{f}_{t,j}(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}_i} \frac{\partial \mathbf{f}_t(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^\top} + \\ & \frac{\partial \tilde{\mathbf{g}}_T(\mathbf{f}_t(\bar{\boldsymbol{\theta}}^*), \bar{\boldsymbol{\theta}}^*)}{\partial \mathbf{f}^\top} \frac{\partial^2 \mathbf{f}_t(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^\top \partial \boldsymbol{\theta}_i}. \end{aligned} \quad (\text{A.12})$$

Using eqs. (A.10) to (A.12), we can now replace all initialized time-varying parameters $\mathbf{f}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$ in eqs. (A.7) to (A.8) by their stationary and ergodic counterparts $\mathbf{f}_t(\boldsymbol{\theta})$ and follow the standard line of argument for proving the asymptotic normality of GMM estimators under Assumption B; see for instance White (1996). \square

Appendix B: Time-varying linear regression models with endogenous covariates

Model

Linear regression models are standard tools in economic analysis. A typical concern is the endogeneity of one of the regressors with the error term due to omitted variable bias or measurement error. Consider the standard linear regression equation

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta}_t + \varepsilon_t,$$

with $E_{t-1}[\mathbf{x}_t^\top \boldsymbol{\beta}_t] \neq 0$. Given the availability of an instrument variable \mathbf{z}_t , the obvious way to estimate $\boldsymbol{\beta}_t$ is via the conditional moment condition

$$E_{t-1}[\mathbf{z}_t(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}_t)] = 0. \quad (\text{B.1})$$

The complication here is that the parameters $\boldsymbol{\beta}_t$ are time-varying. The moment condition eq. (B.1), however, lends itself directly to the GaMM framework by setting $\mathbf{f}_t = \boldsymbol{\beta}_t$. Assuming \mathbf{z}_t and \mathbf{x}_t to be stationary and ergodic, we estimate $\boldsymbol{\Omega}_{t-1} = E[\mathbf{z}_t \mathbf{z}_t^\top]^{-1}$ and $\bar{\mathbf{G}}_t = E[\mathbf{z}_t \mathbf{x}_t^\top]$ by their sample averages.

Simulation results

Consider a simple linear model where both $\beta_{0,t}$ and $\beta_{1,t}$ are time-varying:

$$y_t = \beta_{0,t} + \beta_{1,t} x_t + \varepsilon_{yt}, \quad \varepsilon_{yt} \sim N(0, \sigma_{\varepsilon_y}^2), \quad E[\varepsilon_{yt} \varepsilon_{xt}] = \rho, \quad (\text{B.2})$$

$$x_t = \pi_{0,x} + \pi_{1,x} z_t + \varepsilon_{xt}, \quad \varepsilon_{xt} \sim N(0, \sigma_{\varepsilon_x}^2), \quad z_t \sim N(\mu_z, \sigma_z^2), \quad (\text{B.3})$$

$$\beta_{i,t+1} = \pi_\beta \beta_{i,t} + \eta_{i,t}, \quad \eta_{i,t} \sim N(0, \sigma_{i,\eta}^2), \quad i = 0, 1. \quad (\text{B.4})$$

If $\rho \neq 0$ and $\sigma_{\varepsilon_x}^2 > 0$, we have a standard endogenous regressor problem. For lower values of $\sigma_{\varepsilon_x}^2$, the size of the bias when estimating $\beta_{1,t}$ using standard least-squares based methods, is larger. Define $\mathbf{x}_t^\top = (1, x_t)$, $\mathbf{z}_t^\top = (1, z_t)$, and $\boldsymbol{\beta}_t^\top = (\beta_{0,t}, \beta_{1,t})$, and let $\boldsymbol{\Pi}_x$ be such that $\mathbf{x}_t^\top = \mathbf{z}_t^\top \boldsymbol{\Pi}_x + (0, \varepsilon_{xt})$. As a benchmark model we use an ordinary state space model consisting of equations eq. (B.2) and eq. (B.4) and estimated using standard Kalman Filter methods. This is close to comparing the performance of an instrumental variables (IV) estimator with an ordinary least squares (OLS) estimator in a static context, where in our dynamic context the Kalman Filter and the GaMM estimator take the roles of the OLS and IV estimators, respectively.

Our data generating process uses eqs. (B.2) to (B.4). The parameters we selected are $\rho = 0.5$, $\sigma_{\varepsilon_y}^2 = 1$, $\pi_x = 0.5$, $\mu_z = 0$, $\sigma_z^2 = 1$, and $\boldsymbol{\pi}_\beta^\top = [0.98, 0.98]^\top$. We vary σ_{ε_x} and σ_η to study the effect of different magnitudes of the endogeneity problem and of the time-variation of the coefficients. We use $\sigma_{\varepsilon_x}^2 = \{0.5, 4\}$ and $\sigma_\eta^2 = \{0.01, 0.25, 0.75\}$, where lower values of σ_{ε_x} result in larger biases of the least-squares based estimator. All simulations are repeated for $T = \{1000, 2500, 5000\}$ observations, with 1000 observations corresponding to approximately 4 years of daily data. We obtained similar results for simulations conducted with shorter time-series $T = \{250, 500\}$, as well as for other time-varying patterns for $\boldsymbol{\beta}_t$ than in eq. (B.4), including structural breaks and slowly varying sinusoid waves.

The results of 10,000 replications are presented in Figure B.1 and Figure B.2. The figures summarize a considerable amount of information in a dense way. In the top panel (Panel A), we summarize the results for $\hat{\beta}_{0,t}$, while the bottom panel (Panel B) presents the results for the coefficient $\beta_{1,t}$ corresponding to the endogenous regressor. For each simulation, we summarize the average bias and root mean squared error (RMSE) of the time-varying parameters into a single summary statistic,

$$\text{Average Bias}_i = \frac{1}{T} \sum_{t=1}^T (\hat{\beta}_{i,t} - \beta_{i,t}^{true}), \quad \text{Average RMSE}_i = \left[\frac{1}{T} \sum_{t=1}^T (\hat{\beta}_{i,t} - \beta_{i,t}^{true})^2 \right]^{\frac{1}{2}}. \quad (\text{B.5})$$

The results of the 10,000 simulations are presented as box-plots. In the left-half of each plot, the endogeneity problem is moderate with a high value of $\sigma_{\varepsilon_x}^2 = 4.0$. In the right-hand half of each graph, the endogeneity problem is more severe with $\sigma_{\varepsilon_x}^2 = 0.5$. This becomes particularly important for the coefficient $\beta_{1,t}$ in Panel B. For both the moderate and the high endogeneity bias case, we

B.1

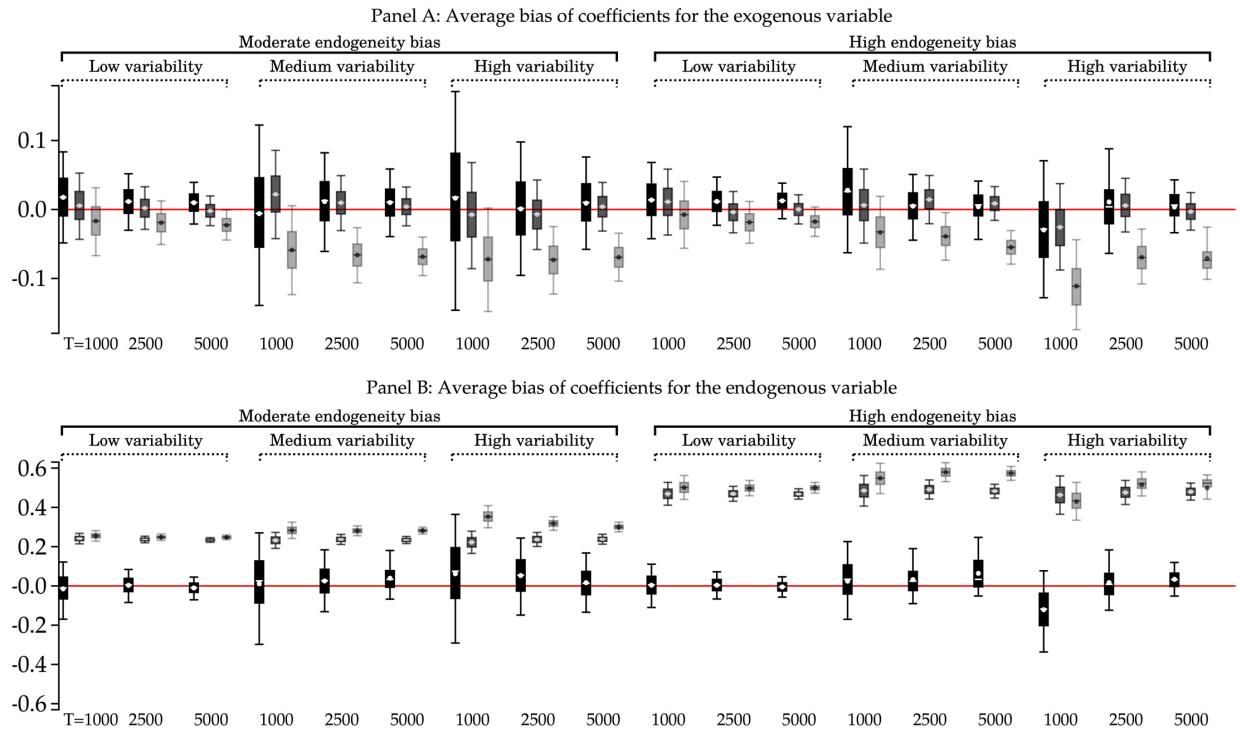
Bias of GaMM for a Simple Endogeneity Problem

The figure compares performance of GaMM (black) to Kalman Filter with AR1 (light grey) or Random Walk (dark grey) dynamics for coefficients. We simulate observations and true parameters from a simple linear model where all coefficients are time-varying:

$$\begin{aligned} y_t &= \beta_{0,t} + x_t \beta_{1,t} + \varepsilon_t \\ x_t &= 0.5z_t + \phi_t \\ \beta_{i,t+1} &= 0.98\beta_{i,t} + \eta_{i,t} \end{aligned}$$

$$E[\varepsilon_t \phi_t] = 0.5 \quad \varepsilon_t \sim N(0, 1), \quad \phi_t \sim N(0, \sigma_\phi) \quad z_t \sim N(0, 1) \quad \eta_{i,t} \sim N(0, \sigma_\eta)$$

We consider a moderate ($\sigma_{\varepsilon_x}^2 = 4.0$; left-hand half of the figure) and a high ($\sigma_{\varepsilon_x}^2 = 0.5$; right-hand half of the figure) degree of the endogeneity problem. For both the moderate and the high endogeneity bias case, we consider different degrees of time-variation in β_t , low ($\sigma_\eta^2 = 0.01$), medium ($\sigma_\eta^2 = 0.25$), and high ($\sigma_\eta^2 = 0.75$). For each of the $2 \times 9 = 18$ combinations, we plot the results for three different simulated sample sizes $T = \{1000, 2500, 5000\}$. Each group of three box-plots corresponds to a combination of sample size, degree of time-variation in β_t , and severity of the endogeneity problem. The three box-plots correspond to three different models: GaMM(1,1) (black), state space model with random walk dynamics (dark grey), and state space model with autoregressive (AR) dynamics of order 1 (light grey). The results are based on 10,000 replications.



consider different degrees of time-variation in β_t , from low ($\sigma_\eta^2 = 0.01$) to high ($\sigma_\eta^2 = 0.75$). For each of the $2 \times 9 = 18$ combinations, we plot the results for three different simulated sample sizes $T = \{1000, 2500, 5000\}$. Each group of three box-plots corresponds to a combination of sample size, degree of time-variation in β_t , and severity of the endogeneity problem. The three box-plots correspond to three different models: GaMM(1,1) (black), state space model with random walk dynamics (dark grey), and state space model with autoregressive (AR) dynamics of order 1 (light grey).

Figure B.1 shows the in-sample performance in terms of the average bias. Typically, out of the two state space models, the specification with random walk dynamics for the parameters performs better and we continue with this model as our main benchmark.

In Panel A, we see that in most cases both the GaMM and the Kalman Filter estimates based on random walk dynamics offer a similar bias performance for the parameter of the exogenous regressor. The average and median biases are close to zero. It is also clear that the distribution of the average bias for the GaMM approach has a higher spread. The relative differences in performance diminish substantially as the sample size increases. For the largest sample sizes the results produced with GaMM are often as accurate as the ones produced with the Kalman Filter.

In Panel B of Figure B.1 we see that GaMM clearly outperforms the Kalman Filter in terms of bias of the estimator for the parameter corresponding to the endogenous variable. Regardless of sample size, size of the time-variability, or magnitude of the endogeneity bias; the Kalman Filter estimates are clearly biased (as expected), whereas the GaMM approach results in unbiased estimates of the parameter path in almost all cases. As expected, the bias of the Kalman Filter estimates is larger for $\sigma_{\varepsilon_x}^2 = 0.5$ than for $\sigma_{\varepsilon_x}^2 = 4.0$. In fact, the 5% of best results generated by the Kalman Filter are in most cases worse than 95% of the results produced by GaMM. As before, the performance of GaMM improves with the sample size, making the improvement over the biased least-squares based methods even more apparent.

Figure B.2 focuses on average root-mean square errors produced by both methods. For both the exogenous and endogenous variables, RMSEs produced by either method are comparable on average. However, the distribution of RMSE produced by GaMM has a considerably heavier right tail: in some cases, the RMSE behaviour of GaMM can be substantially worse than that of the Kalman Filter methods. The number of simulations with poor RMSE produced by GaMM reduces substantially as the sample size increases.

B.2

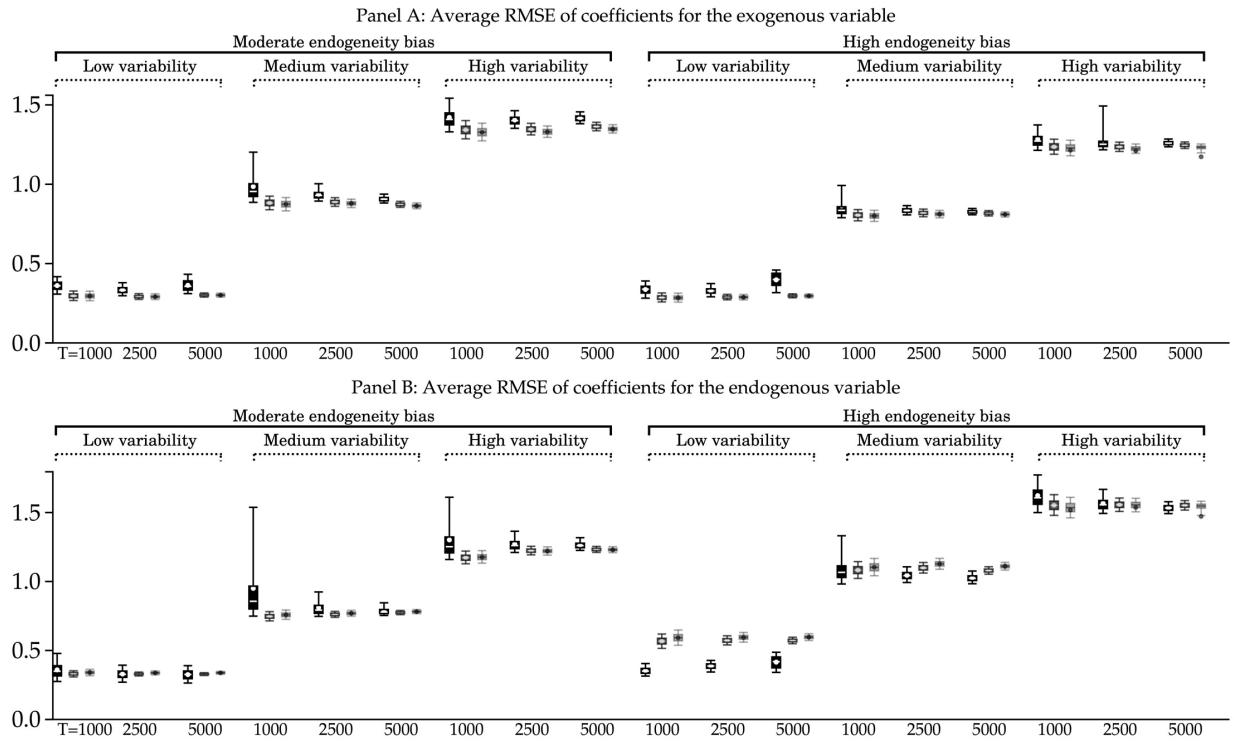
RMSE of GaMM for a Simple Endogeneity Problem

The figure compares performance of GaMM (black) to Kalman Filter with AR1 (light grey) or Random Walk (dark grey) dynamics for coefficients. We simulate observations and true parameters from a simple linear model where all coefficients are time-varying:

$$\begin{aligned} y_t &= \beta_{0,t} + x_t \beta_{1,t} + \varepsilon_t \\ x_t &= 0.5 z_t + \phi_t \\ \beta_{i,t+1} &= 0.98 \beta_{i,t} + \eta_{i,t} \end{aligned}$$

$$E[\varepsilon_t \phi_t] = 0.5 \quad \varepsilon_t \sim N(0, 1), \quad \phi_t \sim N(0, \sigma_\phi) \quad z_t \sim N(0, 1) \quad \eta_{i,t} \sim N(0, \sigma_\eta)$$

We consider a moderate ($\sigma_{\varepsilon_x}^2 = 4.0$; left-hand half of the figure) and a high ($\sigma_{\varepsilon_x}^2 = 0.5$; right-hand half of the figure) degree of the endogeneity problem. For both the moderate and the high endogeneity bias case, we consider different degrees of time-variation in β_t , low ($\sigma_\eta^2 = 0.01$), medium ($\sigma_\eta^2 = 0.25$), and high ($\sigma_\eta^2 = 0.75$). For each of the $2 \times 9 = 18$ combinations, we plot the results for three different simulated sample sizes $T = \{1000, 2500, 5000\}$. Each group of three box-plots corresponds to a combination of sample size, degree of time-variation in β_t , and severity of the endogeneity problem. The three box-plots correspond to three different models: GaMM(1,1) (black), state space model with random walk dynamics (dark grey), and state space model with autoregressive (AR) dynamics of order 1 (light grey). The results are based on 10,000 replications.



Summarizing, the trade-off between the Kalman Filter and GaMM approach seems to mirror in the dynamic parameter case the differences between OLS and IV estimation from the static parameter case. The Kalman Filter produces results that are biased but with low sampling variability, whereas paths estimated by GaMM appear to be unbiased, but at the cost of a higher sampling variance.