TI 2016-067/IV
Tinbergen Institute Discussion Paper

Accounting for Missing Values in Score-Driven Time-Varying Parameter Models

André Lucas
Anne Opschoor
Julia Schaumburg

Faculty of Economics and Business Administration, VU University Amsterdam, and 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.

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)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Accounting for Missing Values in Score-Driven Time-Varying Parameter Models☆

André Lucas[a], Anne Opschoor[a], Julia Schaumburg[a]

[a] *Vrije Universiteit Amsterdam and Tinbergen Institute*

Abstract

We show that two alternative perspectives on how to deal with missing data in the context of the score-driven time-varying parameter models of Creal et al. (2013) and Harvey (2013) lead to precisely the same dynamic transition equations. As score-driven models encompass a wide variety of time-varying parameter models (including generalized autoregressive conditional volatility (GARCH) and duration (ACD) models), the results apply to a wide range of empirically relevant models as applied in economics and statistics.

Keywords: generalized autoregressive score models, missing completely at random, Expectation-Maximization.
JEL: C53, C52

1. Introduction

We address the issue of missing values in observation-driven time-varying parameter models. Such models are widely applied in the empirical economics and econometrics literature and include well-known examples such as the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986), the autoregressive conditional duration (ACD) model of Engle and Russell (1998), the multiplicative error model (MEM) of Engle and Gallo (2006), and many more. In this paper, we focus on the class of score-driven models as introduced and popularized by Creal et al. (2011, 2013) and Harvey (2013). Score-driven models encompass earlier well-known models such as the normal GARCH and ACD models, but also give rise to entirely new models, such as the mixed measurement dynamic factor model of Creal et al. (2014) for macro and credit cycles, the dynamic Nelson-Siegel model with time-varying shape parameter of Quaedvlieg and Schotman (2016) for long-term pension liability hedging, the time-varying copula model of Oh and Patton (2016) to model credit risk in high dimensions, and the matrix-F dynamic model for multivariate realized covariance matrices of Opschoor et al. (2014).[1]

A key property of observation-driven models in the classification of Cox (1981) is that they define the time-varying parameter, such as a mean or a volatility parameter, in terms of its own lags and a (possibly nonlinear) function of lagged observations. This directly poses a problem for updating the time-varying parameter if a missing value is encountered: if the observation is missing, the function of lagged variables cannot be evaluated and the dynamic parameter cannot be updated.

In this paper we discuss how missing values can be dealt with in the context of score-driven models. In particular, we show how score-driven models provide a natural way to include the non-missing part of the data into the updating mechanism, while accounting for the missing data in an intuitive and consistent way. We do so by discussing two alternative perspectives of the missing value problem in our model setting and by showing that the two perspectives lead to exactly the same solution. Under the first perspective, we exploit

☆Schaumburg thanks the Dutch National Science Foundation (NWO, grant VENI451-15-022) for financial support.
Email addresses: `a.lucas@vu.nl` (André Lucas), `a.opschoor@vu.nl` (Anne Opschoor), `j.schaumburg@vu.nl` (Julia Schaumburg)

[1] We refer to www.gasmodel.com for a compendium of papers employing the score-driven approach to time-varying parameter models.

the special feature of score-driven models that the direction in which the dynamic parameter is updated, equals the derivative of the log of the conditional predictive density. If (part of) the data are missing, this predictive density simplifies to a marginal density for the observed data part and the propagation mechanism for the time-varying parameter adapts automatically. This approach is for example used in Creal et al. (2014), where no further motivation was provided. A second perspective sets the missing value problem in the much wider statistical literature about the treatment of missing values; see for example Little and Rubin (2014) for a textbook review. This is also the typical context where the Expectation-Maximization (EM) approach of Dempster et al. (1977) is used for estimation rather than the classical likelihood framework. Though natural, there is as of yet no theory linking the statistical missing value literature to the updating mechanism in score-driven models.

We have two main contributions in this paper. First, we show how a score-driven approach may be devised outside the familiar conditional predictive density context of Creal et al. (2011, 2013). In particular, we use the EM criterion function and its scores to drive the time-varying parameter. This is a novel perspective on the usefulness of score-driven dynamic parameter models. Second, we show that for data that are missing completely at random, a score-driven approach based on the marginal predictive density produces exactly the same result as a score-driven approach based on the EM criterion. This holds irrespective of the scaling choice for the score. This result is an important step in motivating how to deal with missing values in the context of score-driven models as in, for instance, Creal et al. (2014). At the same time, it closes the gap between the literature on score-driven time series models and the classical statistical missing value and EM literature.

The rest of this paper proceeds as follows. Section 2 explains the issue of missing values in the context of the generalized autoregressive score (GAS) model of Creal et al. (2013).[2] Section 3 sets up a score based approach in the EM set-up of Dempster et al. (1977) and proves our theoretical result.

2. Missing-values and standard GAS models

Let x_t and y_t denote two vectors of observations, characterized by the conditional density

$$(x_t, y_t) \sim p(x_t, y_t \mid f_t),$$

where f_t is a time-varying parameter. For example, f_t may contain the conditional means of x_t and y_t, their conditional variances, their correlations, or other higher order features of the conditional distribution; see for example Creal et al. (2011, 2013) and Harvey and Luati (2014) for a range of applications. We consider x_t and y_t as scalars in the remainder of this paper, but all derivations also go through in a higher dimensional context. The conditioning set of $p(\cdot \mid f_t)$ can be augmented with other predetermined information, such as further lags of x_t and y_t. We assume that the dynamics for f_t are given by the generalized autoregressive score dynamics

$$f_{t+1} = \omega + \beta\, f_t + \alpha\, s_t, \qquad s_t = s_t^{\text{joint}} = S_t\, \frac{\partial \log p(x_t, y_t \mid f_t)}{\partial f_t}, \tag{1}$$

where ω, α, and β are parameters that need to be estimated, and S_t is a scaling matrix that may depend on f_t itself. The score step in equation (1) increases the local model fit in the steepest ascent direction of the local likelihood contribution. Blasques et al. (2015) show that a score-driven improvement is locally optimal from an information theoretic perspective: it locally improves the Kullback-Leibler divergence between the unkown true data generating process and the statistical model upon each parameter update. This result holds under quite general conditions, even in cases where the density $p(x_t, y_t \mid f_t)$ is severely mis-specified. Putting S_t equal to the inverse conditional expectation of the squared score,

$$S_t = \mathrm{E}\left[\frac{\partial \log p(x_t, y_t \mid f_t)}{\partial f_t}\, \frac{\partial \log p(x_t, y_t \mid f_t)}{\partial f_t'} \right]^{-1}$$

[2] See also Harvey (2013), who refers to these models as dynamic conditional score models. We also refer to the site gasmodel.com for a much more complete compendium of published and unpublished work on score-driven (GAS) models

corrects the steepest ascent step for the local curvature and creates a Newton-Raphson type improvement step. Other choices for the scaling matrix S_t are also possible.

An important advantage of the score driven approach is that it is observation-driven in the classification of Cox (1981). This enables us to write down the likelihood function $\mathcal{L}_T$ in analytic form via a prediction error decomposition,

$$\mathcal{L}_T = \sum_{t=1}^{T} \log p(x_t, y_t \mid f_t),$$

and to estimate the model's static parameter by standard maximum likelihood estimation.

A drawback of the observation-driven specification is that a problem occurs is y_t (and/or x_t) is missing. Also the score-based framework, being observation-driven, faces this challenge. If y_t is missing, the score s_t cannot be computed, as it depends on both observations. Creal et al. (2014) solve this problem by arguing that missing values y_t are no problem for the score-based approach: at time t, the score should be taken of the conditional observation density. If y_t is missing, the observation density at time t collapses to the marginal density for x_t and thus the score should be computed based on the density of the remaining observation x_t, i.e., using $\partial \log p(x_t|f_t)/\partial f_t$, where $p(x_t|f_t)$ is the marginal conditional density of x_t. The reasoning of Creal et al. (2014) appears intuitive: if y_t is missing, the only information about f_t can be obtained from x_t and therefore from its marginal conditional density, such that

$$s_t = s_t^{\text{marg}} = \mathrm{E}\left[\frac{\partial \log p(x_t \mid f_t)}{\partial f_t} \frac{\partial \log p(x_t \mid f_t)}{\partial f_t'} \right]^{-1} \frac{\partial \log p(x_t \mid f_t)}{\partial f_t}. \tag{2}$$

A different line of argument, however, argues from a conditional expectations perspective.[3] The reasoning is as follows. The above approach of Creal et al. (2014), though intuitive, is rather detached from the common approach of dealing with missing values; see Dempster et al. (1977). Consider for example a setting of a bivariate normal (x_t, y_t) with time-varying means. If we know that x_t and y_t are strongly positively correlated, and if at time t the variable x_t is above its mean while y_t is missing, we might infer that y_t is also above its mean at time t with high probability. Why not use this inferred information to also update the mean of y_t upward?

At first sight, this alternative approach appears to be in line with a typical EM perspective, where we would also work with a conditional expectation (of the log likelihood function) with respect to the missing observations, conditional on the non-missing observations. It contrasts, however, with the previously described approach based on the score of the marginal distribution of x_t. There the mean of y_t would not receive an update signal as the score of the marginal density $p(x_t \mid f_t)$ with respect to the mean of y_t would be zero, and consequently the time-varying mean of y_t would mean-revert via the transition equations (1) and (2).

3. EM score models and equivalence result

To resolve which of the two perspectives discussed in Section 2 is correct, we propose to close the gap between the score approach based on marginal distributions and the EM perspective of dealing with missing values. In doing so, we show that the second line of argument in Section 2 is actually false as the conditional expectation that is taken in the EM algorithm relates to the complete data log likelihood function, and not to y_t itself. We proceed as follows. First, we define a score-based approach based on the local EM type objective function

$$\log \tilde{p}(x_t \mid f_t) := \mathrm{E}\left[\log p(x_t, y_t \mid f_t) \mid x_t, f_t\right]. \tag{3}$$

This is a novel perspective on score-based time-varying parameter modeling, where usually the score is taken of a predictive density in a likelihood framework. Only Creal et al. (2016) is related to our approach in

[3]We found this argument repeatedly popping up when discussing the issue of missing values in score-driven models.

that it gives a score-based generalization of the Generalized Method of Moments (GMM) based on a local moments criterion function. Just as $p(x_t \mid f_t)$ is the local (marginal) likelihood contribution in a set-up based on predictive densities, $\tilde{p}(x_t \mid f_t)$ is the local contribution to the criterion function in an EM set-up. In fact, if there are no missing data, the two approaches coincide and $p(\,\cdot \mid f_t) = \tilde{p}(\,\cdot \mid f_t)$.

We now generalize the standard score-driven transition equation (1) by replacing the scaled derivative of the marginal data likelihood by the scaled derivative of the local conditionally expected complete likelihood function, i.e., the conditional expectation of the log likelihood contribution at time t of both the missing (y_t) and non-missing (x_t) data. Using the EM-based score, we improve the value of f_t by exploiting the full dependence structure between x_t and y_t as given by the joint density. Therefore, if x_t and y_t are highly correlated, this information is also taken into account. We have

$$s_t = s_t^{\mathrm{EM}} = \mathrm{E}\left[\, \frac{\partial \log \tilde{p}(x_t \mid f_t)}{\partial f_t} \; \frac{\partial \log \tilde{p}(x_t \mid f_t)}{\partial f_t'} \,\right]^{-1} \; \frac{\partial \log \tilde{p}(x_t \mid f_t)}{\partial f_t}. \tag{4}$$

We note that this new score type model model automatically deals with missing values in a way that is fully compatible with the EM framework.

The key question now is what differences there are between a transition equation for f_t that is based on score steps from a likelihood perspective and marginal distributions (s_t^{marg}) and one that is based on the EM perspective (s_t^{EM}). The key theoretical result of this paper is that there is *no* difference and that the two approaches are actually fully equivalent. We state this in the following theorem.

Theorem 1. *If the density $p(x_t, y_t \mid f_t)$ is correctly specified, $s_t^{EM} = s_t^{marg}$.*

Proof. We note that

$$\frac{\partial \log p(x_t, y_t \mid f_t)}{\partial f_t} = \frac{\partial\big(\log p(y_t \mid x_t, f_t) + \log p(x_t \mid f_t)\big)}{\partial f_t},$$

such that

$$\frac{\partial \log \tilde{p}(x_t \mid f_t)}{\partial f_t} = \mathrm{E}\left[\frac{\partial \log p(y_t \mid x_t, f_t)}{\partial f_t} \,\middle|\, x_t, f_t\right] + \frac{\partial \log p(x_t \mid f_t)}{\partial f_t} = \frac{\partial \log p(x_t \mid f_t)}{\partial f_t},$$

where the last equality follows directly from the property that the (conditional) score of a correctly specified density has expectation zero. □

The result in Theorem 1 substantiates the way missing values are dealt with in papers such as Creal et al. (2014) and others. At the same time, it ties the score-driven modeling approach in the presence of missing value closer to the older and larger literature on how to deal with missing values via conditional expectation-maximization. Interestingly, it also brings the score-driven approach closer to how missing values are dealt with in the state-space framework. Also there, the marginal score of the density for the non-missing data with respect to the time-varying parameter is key; see for example Durbin and Koopman (2012). Theorem 1 illustrates that this is fully in line with the score-based approach, and that both approaches are moreover strongly related to the EM approach thanks to the particular features of the score as a propagation mechanism for f_t.

References

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*(3), 307–327.

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

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*, 552–563.

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., S. J. Koopman, A. Lucas, and M. J. Zamojski (2016). Generalized autoregressive methods of moments. *Working paper*.

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.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, 1–38.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods.* Number 38. Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica 50*, 987–1008.

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*(1), 3–27.

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

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: with Applications to Financial and Economic Time Series.* Econometric Society Monographs. Cambridge: Cambridge University Press.

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

Little, R. J. and D. B. Rubin (2014). *Statistical analysis with missing data.* John Wiley & Sons.

Oh, D. H. and A. J. Patton (2016). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Journal of Business & Economic Statistics*, to appear.

Opschoor, A., P. Janus, A. Lucas, and D. J. Van Dijk (2014). New heavy models for fat-tailed returns and realized covariance kernels.

Quaedvlieg, R. and P. C. Schotman (2016). Score-driven nelson siegel: Hedging long-term liabilities. *Available at SSRN*.