Maximum Likelihood Estimation of Regression Effects in State Space Models

Hang Qian*
The MathWorks, Inc.

January 15, 2025

Abstract

Unknown parameters, including regression coefficients, in state space models can be estimated by maximum likelihood. An alternative approach is to augment the state vector to include regression coefficients. However, the state estimator obtained by the Kalman filter is numerically different from the maximum likelihood estimator. We address the discrepancy by a novel method based on proper distributions returned by the ordinary Kalman filter without dependency on diffuse initialization. We prove that maximizing a low-dimensional objective function that combines the likelihood, the filtering mean and variance can reproduce the high-dimensional maximum likelihood results.

Keywords: Bayesian analysis, Diffuse prior, Posterior distribution, Likelihood concentration
MSC subject classifications: 62F10, 62L12, 65Y04

*55 Centre Street, Natick, MA, U.S., 01760. Email: matlabist@gmail.com.

1 Introduction

Regression coefficients β in the state space model are time-invariant parameters for maximum likelihood estimation (MLE). Durbin and Koopman (2012, p.148) suggest inclusion of coefficients in the state vector such that $\beta_t = \beta_{t-1} = \beta$, $\forall t$. The Kalman filter provides a closed-form state estimation for β, as an alternative to parameter estimation by numerical methods.

However, the state estimator is numerically different from the MLE result. The discrepancy is acknowledged in the exact initialization literature. Durbin and Koopman (2012, p.175) comment that "[w]hen regression effects are present in the state space model, similar adjustments as for the diffuse loglikelihood function are required". It is possible to make adjustments by alternative definitions of the likelihood function, such as the profile, diffuse or marginal likelihood discussed in Francke et al. (2010), where regression estimation is handled in state space models under diffuse initial conditions. Diffuse initialization techniques are described in Ansley and Kohn (1985), De Jong (1988, 1991), Koopman (1997), Koopman and Durbin (2003), among others.

We propose a novel method that resolves the discrepancy by the ordinary Kalman filter, without dependency on diffuse state initialization. Exact MLE results can be reproduced even if the initial state variance is small. The novelty is an objective function that combines the likelihood, the filtering mean and variance returned by the ordinary Kalman filter upon completion of the forward recursion. We prove that the maximizer of the low-dimensional objective function exactly replicates the high-dimensional MLE results. The strength of our method is to avoid improper distributions and divergent objective functions.

The remainder of the paper is organized as follows. Section 2 contrasts direct MLE versus inclusion of regression coefficients in the state vector and shows the discrepancy. Section 3 proves the exact MLE by post-processing the Kalman filter output. Section 4 is devoted to an application of the panel data regression in the state space framework. Section 5 concludes the paper.

2 State and Parameter Estimation Discrepancy

Consider a normal linear state space model with regression effects in the state and/or observation equations:

$$Y_t = Z_t \alpha_t + X_t \dot{\beta} + \varepsilon_t,$$

$$\alpha_t = T_t \alpha_{t-1} + W_t b + R_t \eta_t,$$

where the disturbances ε_t and η_t follow normal distributions $N(0, H_t)$ and $N(0, Q_t)$, respectively. Here α_t is an unobserved state vector and $\beta = \begin{pmatrix} \dot{\beta} \\ b \end{pmatrix}$ is a time-invariant parameter vector for regressors X_t and W_t. The distinction between states and parameters is crucial in state space modeling. Let us also define θ as a set of nuisance parameters (including variance components) that determine Z_t, T_t, R_t, H_t, Q_t.

The recursive prediction and update steps of the Kalman filter yield the likelihood function $p(Y|\beta, \theta)$ in the prediction error decomposition form (Durbin and Koopman, 2012, p.171). Ideally, unknown parameters β and θ can be estimated by MLE:

$$\begin{pmatrix} \hat{\beta} \\ \hat{\theta} \end{pmatrix} = \arg\max_{\beta,\theta} p(Y|\beta, \theta). \tag{1}$$

Equation (1) does not have a closed-form maximizer. In practice, we specify the log likelihood $\ln p(Y|\beta, \theta)$ as the objective function for numerical optimization. As an alternative to numerical differentiation, Harvey (1990), Hooker (1994), Koopman and Shephard (1992) and Nagakura (2021) discuss the exact gradient of the log likelihood. Koopman and Shephard (1992) conclude that the gradient associated with system matrices Z_t and T_t can be better evaluated numerically than analytically. Also, it is possible to concentrate out a scale parameter and reduce the dimensionality of the estimation problem by one parameter (Durbin and Koopman, 2012, p.36).

Direct MLE based on Equation (1) is practically unattractive in the presence of many predictors such as a large set of dummy variables. For example, we could add 12 monthly dummy variables to address seasonality. High-dimensional numerical optimization is computationally intensive and could terminate at local maxima or saddle points. We view the MLE estimators $\hat{\beta}, \hat{\theta}$ as the benchmark for the purpose of comparison with alternative estimators.

An alternative approach, common in practice, is to augment the state vector to include β itself. This way, the Kalman filter provides state estimation for β instead of parameter estimation. As suggested by Durbin and Koopman (2012, p.148) and Commandeur et al. (2011), time-invariant parameters are endowed with a dynamic structure in an enlarged state space model:

$$Y_t = \begin{pmatrix} Z_t & X_t & 0 \end{pmatrix} \begin{pmatrix} \alpha_t \\ \dot{\beta}_t \\ b_t \end{pmatrix} + \varepsilon_t,$$

$$\begin{pmatrix} \alpha_t \\ \dot{\beta}_t \\ b_t \end{pmatrix} = \begin{pmatrix} T_t & 0 & W_t \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} \alpha_{t-1} \\ \dot{\beta}_{t-1} \\ b_{t-1} \end{pmatrix} + \begin{pmatrix} R_t \\ 0 \\ 0 \end{pmatrix} \eta_t.$$

Since the regression components have been incorporated in the state vector, it seems reasonable to estimate the remaining parameters θ by solving a smaller-scale optimization problem:

$$\tilde{\theta} = \arg\max_{\theta} p\left(Y \left| \theta \right.\right), \tag{2}$$

where $p\left(Y \left| \theta \right.\right)$ is the likelihood function of the enlarged state space model. The objective functions in Equations (1) and (2) are related by

$$p\left(Y \left| \theta \right.\right) = \int p\left(\beta \left| \theta \right.\right) p\left(Y \left| \beta, \theta \right.\right) d\beta,$$

where a normal prior $p\left(\beta \left| \theta \right.\right)$ is specified for the initial state of β. The conventional wisdom is diffuse initialization. The main contribution of the paper is to show applicability of the proper prior, even with a small variance. We only use the ordinary Kalman filter throughout the paper.

The ordinary Kalman filter completes the forward recursion by returning at least three variables: the mean of the last-period state filtering distribution (posterior mean) $\mu(\theta) = E\left(\beta \left| Y, \theta \right.\right)$, the variance of the last-period state filtering distribution (posterior variance) $\Omega(\theta) = Var\left(\beta \left| Y, \theta \right.\right)$ and the likelihood $p\left(Y \left| \theta \right.\right)$.

Given the estimated parameters $\tilde{\theta}$ in Equation (2), we take the posterior mean as the state estimator:

$$\tilde{\beta} = \mu\left(\tilde{\theta}\right). \tag{3}$$

Unfortunately, the state estimator $\tilde{\beta}$ in Equation (3) is different from the MLE $\hat{\beta}$ in Equation (1), and $\tilde{\theta} \neq \hat{\theta}$ as well, regardless of the prior specification for $p\left(\beta\left|\theta\right.\right)$.

We demonstrate the discrepancy by a local level model in Commandeur et al. (2011):

$$Z_t = 1, T_t = 1, R_t = \sigma_\eta, H_t = \sigma_\varepsilon^2, Q_t = 1,$$

with 6 regressors in the observation equation. Table 1 compares the state and parameter estimation results by artificial data with sample sizes $n = 50, 100, 500$. We specify a nearly diffuse prior for β: $N\left(0, \kappa I\right)$ with $\kappa = 10^7$. We note that $\hat{\beta}$ and $\tilde{\beta}$ are numerically different in all cases, and the discrepancy is larger under smaller n. Increase of κ cannot reduce the discrepancy. Table 1 also provides a preview of our exact estimator that rectifies Equation (2) by an additional term $\left|\Omega\left(\theta\right)\right|^{-1/2}$ in the objective function. In the next section, we will explain the reason that the additional term is crucial for settling the difference between the state and parameter estimation results.

3 Exact MLE

The aim of this section is to identify the source of discrepancy in Table 1 and show that the discrepancy can be removed if the objective function in Equation (2) is rectified by $\mu\left(\theta\right)$ and $\Omega\left(\theta\right)$, which are readily available as the Kalman filter output without additional computing efforts. Our method has no dependency on diffuse initialization, as we specify a normal prior $N\left(\underline{\mu}, \underline{\Omega}\right)$ for the initial state of β:

$$p\left(\beta\left|\theta\right.\right) = \phi\left(\beta; \underline{\mu}, \underline{\Omega}\right),$$

where $\phi\left(\cdot\right)$ denotes the multivariate normal density function. The main result of the paper is summarized in Proposition 1, which shows the finite-sample exact MLE.

	$n=50$			$n=100$			$n=500$		
	MLE	State	Exact	MLE	State	Exact	MLE	State	Exact
β_1	-2.268	-2.163	-2.268	17.569	17.485	17.568	-1.295	-1.311	-1.295
β_2	0.907	0.681	0.907	-4.521	-4.449	-4.521	-1.632	-1.633	-1.632
β_3	4.050	3.966	4.050	1.795	1.771	1.795	10.489	10.502	10.489
β_4	5.071	5.144	5.071	4.800	4.783	4.800	11.479	11.487	11.478
β_5	-10.727	-10.322	-10.727	4.244	4.186	4.244	-0.420	-0.416	-0.421
β_6	19.041	18.873	19.041	-1.819	-1.726	-1.819	2.153	2.148	2.153
σ_η	1.074	1.077	1.074	1.261	1.274	1.261	0.887	0.888	0.887
σ_ε	1.288	1.425	1.288	1.754	1.828	1.754	2.049	2.064	2.049

Table 1: State and parameter estimation under large variance initialization. Data are simulated with sample sizes 50, 100 and 500. MLE results $\hat{\beta}, \hat{\theta}$ in Equation (1) are compared to state estimators $\tilde{\beta}, \tilde{\theta}$ in Equations (2) and (3), as well as exact estimators β^*, θ^* in Equations (4) and (5). The prior variance of β is 10^7.

Proposition 1 *Under a normal prior $N\left(\underline{\mu}, \underline{\Omega}\right)$ and the Kalman filter output $\mu(\theta)$, $\Omega(\theta)$ and $p(Y|\theta)$, the MLE results in Equation (1) are reproduced by $\hat{\beta} = \beta^*(\theta^*)$ and $\hat{\theta} = \theta^*$, where*

$$\beta^*(\theta) = \left[\Omega^{-1}(\theta) - \underline{\Omega}^{-1}\right]^{-1}\left[\Omega^{-1}(\theta)\,\mu(\theta) - \underline{\Omega}^{-1}\underline{\mu}\right], \tag{4}$$

$$\theta^* = \arg\max_{\theta} \frac{\phi\left[\beta^*(\theta); \mu(\theta), \Omega(\theta)\right] p(Y|\theta)}{\phi\left[\beta^*(\theta); \underline{\mu}, \underline{\Omega}\right]}. \tag{5}$$

Proof: The maximization problem in Equation (1) can be solved by optimizing β conditional on θ, and then optimizing the concentrated likelihood with respect to θ:

$$\beta^*(\theta) = \arg\max_{\beta} p(Y|\beta, \theta),$$

$$\theta^* = \arg\max_{\theta} p(Y|\beta^*(\theta), \theta).$$

By Bayes' rule, the objective function can be written as

$$\beta^*(\theta) = \arg\max_{\beta} \frac{p(\beta|Y, \theta)\, p(Y|\theta)}{p(\beta|\theta)},$$

where the prior $p(\beta\,|\theta)$ and the posterior $p(\beta\,|Y,\theta)$ are normal distributions $N\left(\underline{\mu},\underline{\Omega}\right)$ and $N(\mu(\theta),\Omega(\theta))$, respectively. We have quadratic forms in the normal density functions:

$$\beta^*(\theta) = \arg\max_{\beta} -[\beta-\mu(\theta)]'\,\Omega^{-1}(\theta)\,[\beta-\mu(\theta)] + \left(\beta-\underline{\mu}\right)'\underline{\Omega}^{-1}\left(\beta-\underline{\mu}\right).$$

The quadratic objective function has a unique maximizer:

$$\beta^*(\theta) = \left[\Omega^{-1}(\theta) - \underline{\Omega}^{-1}\right]^{-1}\left[\Omega^{-1}(\theta)\,\mu(\theta) - \underline{\Omega}^{-1}\underline{\mu}\right].$$

Plugging $\beta^*(\theta)$ into the likelihood function, we obtain the concentrated likelihood, of which the maximizer reproduces the MLE results: $\hat{\beta} = \beta^*(\theta^*)$ and $\hat{\theta} = \theta^*$. ■

Proposition 1 is an application of Bayes' rule on normal distributions. The posterior is a compromise between data and prior information, as the posterior mean is a weighted average of the prior mean and the MLE results (Gelman et al., 2014). Because the posterior has already been computed by the ordinary Kalman filter upon completion of the forward recursion, the novelty of Proposition 1 is to solve an inverse problem: infer the MLE results from the given prior and posterior distributions.

Proposition 1 holds under an arbitrary choice of the prior variance $\underline{\Omega}$. Under a proper prior $p(\beta\,|\theta)$ with a small variance $\underline{\Omega}$, both the posterior $p(\beta\,|Y,\theta)$ and the marginal likelihood $p(Y\,|\theta)$ are proper. Consequently, the objective function in Equation (5) is non-divergent, avoiding technical complications associated with improper priors and the diffuse likelihood.

Consider the limiting case that the prior precision matrix tends to zero: $\underline{\Omega}^{-1} \to 0$, which corresponds to a diffuse prior $p(\beta\,|\theta) \propto 1$. In that case, Equations (4) and (5) in Proposition 1 reduce to a surprisingly simple form:

$$\beta^*(\theta) = \mu(\theta), \tag{6}$$

$$\theta^* = \arg\max_{\theta} |\Omega(\theta)|^{-1/2}\, p(Y\,|\theta). \tag{7}$$

The intuition of Equation (6) is that the maximizer of the posterior density under the diffuse prior is also the maximizer of the likelihood function: $\beta^*(\theta) = \arg\max_{\beta} p(\beta\,|Y,\theta)$. For a unimodal and symmetric distribution (e.g., the normal distribution), the maximizer of the posterior density is the same as the posterior mean. It follows that $\beta^*(\theta) = \mu(\theta)$.

Equation (7) is obtained by evaluating the normal posterior density at the mean/mode:

$$\phi\left[\beta^*(\theta);\mu(\theta),\Omega(\theta)\right]\propto\left|\Omega(\theta)\right|^{-1/2},$$

$$\phi\left[\beta^*(\theta);\underline{\mu},\underline{\Omega}\right]\propto 1.$$

Compared to Equation (7), the objective function in Equation (2) is mis-specified in the absence of the term $\left|\Omega(\theta)\right|^{-1/2}$, which is the root cause of the discrepancy between the state and parameter estimation results shown in Table 1.

It is of interest of discuss some special cases that a mis-specified objective function like Equation (2) happens to provide the correct results. First, if the model does not have any nuisance parameters θ, the state estimator on β is the same as parameter estimation results by MLE. Second, if the posterior density function $p(\beta\,|Y,\theta)$ does not depend on θ, Equations (2) and (7) are equivalent for optimization, so that $\tilde{\theta}=\hat{\theta}$ and $\tilde{\beta}=\hat{\beta}$. Third, suppose that $p(\beta\,|Y,\theta)$ is a function of θ, but the maximizer $\beta^*(\theta)$ does not depend on θ. In that case, we have correct estimation results on regression effects, although estimators of other parameters may not be correct: $\tilde{\theta}\neq\hat{\theta}$ and $\tilde{\beta}=\hat{\beta}$. The fixed effect panel data regression in Section 4 demonstrates such a special case.

Table 2 illustrates Proposition 1 by a numerical exercise, in which the regression effects β have small variances in the prior distribution. Tables 1 and 2 use the same model and data, but here we specify $\underline{\mu}=0$ and $\underline{\Omega}=\kappa I$ with $\kappa=0.01$. The small variance corresponds to a tight prior on β with strong shrinkage towards the prior mean. Consequently, the estimators obtained by Equations (6) and (7) are close to zero, while Equations (4) and (5) reproduce the benchmark MLE results.

4 An Application

Our research on regression effects in state space models was motivated by an experiment on the Kalman filter software for reproducing results of the panel data regression

$$Y_{it}=\gamma_i+X_{it}\beta+\sigma_\varepsilon\varepsilon_{it},$$

	$n=50$			$n=100$			$n=500$		
	MLE	Shrinkage	Exact	MLE	Shrinkage	Exact	MLE	Shrinkage	Exact
β_1	-2.268	0.008	-2.268	17.569	0.013	17.569	-1.295	0.050	-1.295
β_2	0.907	0.007	0.907	-4.521	0.011	-4.521	-1.632	0.051	-1.632
β_3	4.050	0.009	4.050	1.795	0.011	1.795	10.489	0.059	10.489
β_4	5.071	0.008	5.071	4.800	0.011	4.800	11.479	0.059	11.478
β_5	-10.727	0.006	-10.727	4.244	0.012	4.244	-0.420	0.051	-0.421
β_6	19.041	0.009	19.041	-1.819	0.013	-1.819	2.153	0.053	2.153
σ_η	1.074	1.175	1.074	1.261	1.364	1.261	0.887	0.982	0.887
σ_ε	1.288	1.670	1.288	1.754	2.149	1.754	2.049	2.369	2.049

Table 2: State and parameter estimation under small variance initialization. Data are simulated with sample sizes 50, 100 and 500. MLE results in Equation (1) are compared to shrinkage estimators by in Equations (6) and (7), as well as exact estimators by Equations (4) and (5). The prior variance of β is 0.01.

for $i = 1, \ldots, m$ and $t = 1, \ldots, n$. The unobserved heterogeneity (as the state variable) γ_i follows $N\left(0, \sigma_\gamma^2\right)$ in the random effects model, while the fixed effects allow for arbitrary dependence between the unobserved heterogeneity and regressors (Wooldridge, 2010). For panel data regressions, the maximizer in Equation (1) can be obtained by the generalized least squares for the random effects and least squares dummy variable method for the fixed effects, so that we can compare estimators $\hat{\beta}, \hat{\theta}$ in Equations (1) with $\tilde{\beta}, \tilde{\theta}$ in Equations (2) and (3), as well as β^*, θ^* in Equations (4) and (5).

We consider a well-known labor market dataset analyzed by Cornwell and Rupert (1988), which is also used extensively in Greene (2008). The panel data contain the wage information of 595 workers in 7 years from 1976 to 1982. To study the returns to schooling, the response variable in the panel data regression is the log wage, and the regressors include years of work experience, weeks worked, occupation, indicators of manufacturing industry, residence in south area and SMSA, under union contract, years of education, and so on. Time invariant regressors are included only in the model for random effects.

Table 3 compares the parameter estimation results for the random effects (first three

	Random Effect Model			Fixed Effect Model		
	MLE	Exact	State	MLE	Exact	State
Exp	0.099	0.099	0.0951	0.1114	0.1114	0.1114
Exp2	-0.0005	-0.0005	-0.0005	-0.0004	-0.0004	-0.0004
Wks	0.0008	0.0008	0.0008	0.0007	0.0007	0.0007
Occ	-0.0209	-0.0209	-0.0219	-0.0192	-0.0192	-0.0192
Ind	0.018	0.018	0.0174	0.0208	0.0208	0.0208
South	0.009	0.009	0.0081	0.0031	0.0031	0.0031
SMSA	-0.0448	-0.0448	-0.0436	-0.0419	-0.0419	-0.0419
MS	-0.0441	-0.0441	-0.0488	-0.0286	-0.0286	-0.0286
Union	0.0348	0.0348	0.0367	0.0295	0.0295	0.0295
Year2	-0.0414	-0.0414	-0.0519	-0.0077	-0.0077	-0.0077
Year3	0.008	0.008	0.0025	0.0256	0.0256	0.0256
Year4	0.0273	0.0273	0.0269	0.0285	0.0285	0.0285
Year5	0.0399	0.0399	0.0447	0.0242	0.0242	0.0242
Year6	0.04	0.04	0.0502	0.0074	0.0074	0.0074
Fem	-0.2045	-0.2045	-0.2269			
Ed	0.1295	0.1295	0.1249			
Blk	-0.2506	-0.2506	-0.2431			
σ_ε^2	0.0237	0.0237	0.0245	0.0196	0.0196	0.0238
σ_γ^2	0.5837	0.5837	0.4218			

Table 3: Panel data regression with random effects (first three columns) and fixed effects (last three columns). MLE results $\hat{\beta}, \hat{\theta}$ in Equation (1) are compared to exact estimators β^*, θ^* in Equations (4) and (5), as well as state estimators $\tilde{\beta}, \tilde{\theta}$ in Equations (2) and (3).

columns) and the fixed effects (last three columns) models. For the model of random effects, the unknown parameters are β and $\theta = \left(\sigma_{\varepsilon}^2, \sigma_{\gamma}^2\right)$. Empirical results confirm that $\tilde{\theta}$ obtained by Equation (2) is different from $\hat{\theta}$ by Equation (1). The discrepancy leads to $\tilde{\beta} \neq \hat{\beta}$. In contrast, θ^* from Equation (5) corrects the bias, so that β^* matches $\hat{\beta}$. For the model of fixed effects, the unknown parameters are β and $\theta = \sigma_{\varepsilon}^2$. The empirical results are interesting: estimators of θ by Equations (1) and (2) are different, but estimators of β are the same. This is a special case discussed in Section 3, where the maximizer of $p(\beta \mid Y, \theta)$ does not depend on θ.

5 Conclusion

In the state space literature, initial states can be viewed as unknown parameters or diffuse random variables (Shephard and Harvey, 1990). However, for classical inference of the regression effects, it is clear that the coefficients β are unknown parameters of interest and the MLE in Equation (1) provides the benchmark for parameter estimation. Inclusion of β in the state vector induces a gap between the state estimation by the Kalman filter and the parameter estimation by MLE. Our method bridges the gap by reassembling the output variables of the ordinary Kalman filter initialized under a proper prior, and the prior variances are not necessarily large. Parsimony is the appeal of the paper, as the minimum requirement of implementing our method is the ordinary Kalman filter that returns the last-period filtering mean, variance, and the likelihood upon completion of the forward recursion.

References

Ansley, C. F. and R. Kohn (1985). Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. *Annals of Statistics 13*, 1286–1316. 2

Commandeur, J., S. J. Koopman, and M. Ooms (2011). Statistical software for state space methods. *Journal of Statistical Software 41*, 1–18. 4, 5

Cornwell, C. and P. Rupert (1988). Efficient estimation with panel data: An empirical comparison of instrumental variable estimators. *Journal of Applied Econometrics 3*, 149–155. 9

De Jong, P. (1988). The likelihood for a state-space model. *Biometrika 75*, 165–169. 2

De Jong, P. (1991). The diffuse Kalman filter. *The Annals of Statistics 19*(2), 1073–1083. 2

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford: Oxford University Press. 2, 3, 4

Francke, M. K., S. J. Koopman, and A. F. de Vos (2010). Likelihood functions for state space models with diffuse initial conditions. *Journal of Time Series Analysis 31*, 407–414. 2

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin (2014). *Bayesian Data Analysis (Third Edition)*. Boca Raton: CRC Press. 7

Greene, W. H. (2008). *Econometric Analysis, Sixth Edition*. New Jersey: Prentice Hall. 9

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press. 3

Hooker, M. A. (1994). Analytic first and second derivatives for the recursive prediction error algorithm's log likelihood function. *IEEE Transactions on Automatic Control 39*(3), 662–664. 3

Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association 92*, 1630–1638. 2

Koopman, S. J. and J. Durbin (2003). Filtering and smoothing of state vector for diffuse state-space models. *Journal of Time Series Analysis 24*(1), 85–98. 2

Koopman, S. J. and N. G. Shephard (1992). Exact score for time series models in state space form. *Biometrika 79*(4), 823–826. 3

Nagakura, D. (2021). Computing exact score vectors for linear Gaussian state space models. *Communications in Statistics - Simulation and Computation 50*(8), 2313–2326. 3

Shephard, N. G. and A. C. Harvey (1990). On the probability of estimating a deterministic component in the local level model. *Journal of Time Series Analysis 11*, 339–347. 11

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data, Second Edition*. Cambridge: The MIT Press. 9