TI 2024-003/III
Tinbergen Institute Discussion Paper

A robust Beveridge-Nelson decomposition using a score-driven approach with an application

Francisco Blasques[1]
Janneke van Brummelen[2]
Paolo Gorgi[3]
Siem Jan Koopman[4]

1 Vrije Universiteit Amsterdam and Tinbergen Institute
2 Vrije Universiteit Amsterdam and Tinbergen Institute
3 Vrije Universiteit Amsterdam and Tinbergen Institute
4 Vrije Universiteit Amsterdam and Tinbergen Institute

A robust Beveridge-Nelson decomposition using a score-driven approach with an application*

F. Blasques[a], J. van Brummelen[a], P. Gorgi[a], and S.J. Koopman[a]

[a]Vrije Universiteit Amsterdam and Tinbergen Institute

Abstract

The equivalence of the Beveridge-Nelson decomposition and the trend-cycle decomposition is well established. In this paper we argue that this equivalence is almost immediate when a Gaussian score-driven location model is considered. We also provide a natural extension towards heavy-tailed distributions for the disturbances which lead to a robust version of the Beveridge-Nelson decomposition.

Keywords: trend and cycle, filtering, autoregressive integrated moving average model, score-driven model, heavy-tailed distributions.
JEL codes: C22, E32.

1 Introduction

The Beveridge-Nelson (BN) decomposition, as developed by Beveridge and Nelson (1981), can be regarded as a model-based method for decomposing time series into permanent and transitory components. When applied to economic output time series such as Gross Domestic Product (GDP) and Industrial Production Index (IPI), the permanent component can be interpreted as the trend while the transitory component is often referred to as the cycle or more specifically as the *business cycle*. The BN decomposition implicitly defines the trend as a conditional long-run expectation,

*This version: December 22, 2023. Corresponding author: J. van Brummelen. F. Blasques thanks the Dutch Research Council (VI.Vidi.195.099) for support.

when all dynamic predictability has died out and all deterministic trend effects have been accounted for. The BN cycle of an economic output variable is a sensible estimate of the *output gap*, provided that the underlying model can produce accurate predictions over short and medium term horizons.

The BN decomposition is closely related to the extraction of trend and cycle components from an unobserved components (UC) time series model such as the ones proposed by Watson (1986), Clark (1987) and Harvey and Jaeger (1993). In particular, Oh et al. (2008) and Morley (2011) have investigated the econometric differences and interrelations between BN and UC decompositions. Anderson et al. (2006) has shown the econometric equivalence of the BN decomposition and the UC decomposition based on a single-source of error model. We build on this contribution but propose a more simplified parametrization, without making a reference to a linear state space representation and without making use of Kalman filtering and smoothing methods (Durbin and Koopman, 2012). Furthermore, we propose to extend the linear formulation to a non-linear version by considering a heavy-tailed distribution for the noise term. This extension is imminent once we have recognized that the BN decomposition can be expressed as a dynamic score-driven location model as introduced by Creal et al. (2013) and Harvey (2013). The foundations of the methods are presented in Blasques et al. (2024), we will discuss the implications for BN in detail in this paper.

2 The ARIMA model and the Beveridge-Nelson decomposition

2.1 ARIMA model and BN decomposition

Let $\{x_t\}$ be generated by an autoregressive integrated moving average (ARIMA) process of order $(p, 1, q)$ with deterministic drift d. The well known BN decomposition breaks x_t into trend (τ_t^*) and cycle (ψ_t^*) components $x_t = \tau_t^* + \psi_t^*$[1]. The BN trend τ_t^* can be interpreted as the permanent component of x_t, as it is exactly the long-run expectation $\lim_{h\to\infty} \mathbb{E}[x_{t+h}|\mathcal{F}_t] - dh = \tau_t^*$ (see also Beveridge and Nelson, 1981), where $\{\mathcal{F}_t\}$ is the filtration generated by the sequence of observations $\{x_t\}$. It follows immediately that ψ_t^* is a transitory component that has long-run expectation equal to zero.

[1] More details are available in the Online Appendix A

2.2 A score-driven representation of BN

By making the distinction between a long-run and a short-run component explicit, the following score-driven unit-root location model for $\{x_t\}$ delivers the BN decomposition:

$$x_t = \mu_t + \varepsilon_t, \qquad \mu_t = \tau_t + \psi_t, \tag{1}$$

where τ_t is the non-stationary trend and ψ_t is the stationary remainder, which we can specify as in a score-driven location model, that is

$$\tau_{t+1} = \omega + \tau_t + \kappa s_t, \qquad \psi_{t+1} = \beta_1 \psi_t + \ldots + \beta_p \psi_{t-p+1} + \alpha_1 s_t + \ldots + \alpha_q s_{t-q+1}, \tag{2}$$

with coefficients ω, κ, $\alpha_1, \ldots, \alpha_q$ and $\beta_1, \ldots, \beta_p$ treated as fixed unknown parameters, and the disturbance ε_t is normally, independent and identically distributed with mean zero and variance σ^2. The roots of the AR polynomial $\beta_p(L) = 1 - \beta_1 L - \cdots - \beta_p L^p$ are assumed to be outside the unit circle. The dynamic model specification (1)-(2) belongs to the class of score-driven models proposed by Creal et al. (2013) and Harvey (2013), so s_t in (2) is the scaled score of the predictive density with respect to the location μ_t. It is well known that in the current setting with Gaussian innovations, we have $s_t \equiv \varepsilon_t$; see Online Appendix B. It is immediately clear from (2) and the current assumptions that $\{\tau_t\}$ is a random walk with drift and $\{\psi_t\}$ is a stationary ARMA(p, q) process.

It is easy to show that the model in (1)-(2) is equivalent to an ARIMA$(p, 1, \max(p, q) + 1)$ model for $\tau_1 = x_1$, see Online Appendix C. Furthermore, the long-run expectation of x_t is equal to

$$\lim_{h\to\infty} \mathbb{E}[x_{t+h}|\mathcal{F}_t] - \omega h = \lim_{h\to\infty} \mathbb{E}[\tau_{t+h} + \psi_{t+h} + \varepsilon_{t+h}|\mathcal{F}_t] - \omega h = \tau_{t+1} - \omega\,,$$

as we can write $\tau_{t+h} = \omega(h-1) + \tau_{t+1} + \sum_{k=1}^{h-1} \kappa \varepsilon_{t+k}$, $\{\psi_t\}$ is stationary and ergodic, and has mean zero by construction, and the elements of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are independent over time. Due to these equivalent representations, it follows that the component τ_t is equal to the long-run component of the BN decomposition at time $t-1$, corrected for the deterministic drift: $\tau_t = \tau^*_{t-1} + \omega$. Hence, the parameter κ is the long-run coefficient in the BN decomposition,[2] just as in the single-source

[2] see Online Appendix A for additional details.

of error model of Anderson et al. (2006). It follows directly that $x_t - (\tau_{t+1} - \omega)$ is the short-term BN component of x_t.

We can conclude that the BN decomposition can conveniently be constructed using the score-driven two-component model in (1)-(2). As this is an observation-driven model, the values of τ_t and ψ_t can straightforwardly be filtered for a given sample $\{x_t\}_{t=1}^{T}$, using the following filtering equations:

$$\hat{\tau}_{t+1} = \omega + \hat{\tau}_t + \kappa(x_t - \hat{\tau}_t - \hat{\psi}_t),$$
$$\hat{\psi}_{t+1} = \beta_1\hat{\psi}_t + \ldots + \beta_p\hat{\psi}_{t-p+1} + \alpha_1(x_t - \hat{\tau}_t - \hat{\psi}_t) + \ldots + \alpha_q(x_{t-q+1} - \hat{\tau}_{t-q+1} - \hat{\psi}_{t-q+1}),$$

where we set $\hat{\tau}_1 = x_1$ and $\hat{\psi}_1 = 0$. When using filters that arise from observation-driven models, it is well known that so-called filter invertibility is a desirable property as it, for example, ensures that the effect of the filter initialization dies out in the limit (Straumann and Mikosch, 2006). The multivariate filter above is invertible whenever the MA polynomial of the equivalent ARIMA model is invertible, but invertibility can also be verified directly using its score-driven specification (Blasques et al., 2024).

Some remarks are in place. First, the score-driven model above can be extended straightforwardly to accommodate ARIMA models with higher orders of integration. For instance, by adding a third component γ_t which is specified as $\gamma_{t+1} = \gamma_t + z_t$ with $z_{t+1} = z_t + \lambda s_t$ for some coefficient λ, the model becomes equivalent to an ARIMA$(p, 2, q)$ model. Second, our approach does not rely on any application of the Kalman filter and smoothing methods (Durbin and Koopman, 2012). The "filtering" equations above suffice for the computation of trend and cycle components. Third, the maximum likelihood estimator (MLE) of the static parameters relies on the numerical maximization of the likelihood function, which can be constructed simply using the filters above.

3 A robust BN decomposition

The Gaussian score-driven model introduced in Section 2 can be modified by adopting different distributions for ε_t in (1). In non-Gaussian settings the scaled score s_t will typically no longer be a linear function of ε_t, which implies that the score-driven model is no longer equivalent to a linear ARIMA model. For example, for Student's t distributed ε_t with ν degrees of freedom and scale

σ^2, the scaled score function becomes robust to outliers,

$$s_t = \frac{\varepsilon_t}{1 + \nu^{-1}\sigma^{-2}\varepsilon_t^2} \, ,$$

and see Harvey and Luati (2014), which for finite ν downweights large values of ε_t. Other examples are a model where the innovations are distributed according to the exponential generalized beta distribution of the second kind (Caivano and Harvey, 2014) or a finite mixture of normal distributions (Blasques et al., 2024). In these cases the model is equivalent to an ARIMA with nonlinear MA terms. Crucially, the decomposition as formulated in (1)-(2) as well as the interpretations of trend τ_t and the stationary cycle ψ_t remain valid. The trend $\tau_t - \omega$ remains to be the long-run expectation at time $t - 1$. As such, the model no longer leads to a regular BN decomposition, but instead it leads to a non-linear equivalent of it, which may for instance be robust to outliers.

Blasques et al. (2024) study the theoretical properties of MLE for models of the form (1)-(2). In particular, consistency and asymptotic normality of the MLE is established for ε_t distributed according to a finite mixture of normals for the case without short-run component ψ_t. Furthermore, no issues seem to arise concerning the theoretical properties of the MLE in the setting with a short-run component. Blasques et al. (2024) discuss in detail that in the unit root setting, it is not possible to establish filter invertibility for models based on distributions with fatter tails than the normal distribution, say the Student's t distribution. Hence, consistency and asymptotic normality of the MLE cannot be established for the Student's t, while it can be established for the mixture of normals. This consideration could be a motivation to adopt the mixture of normals in model (1)-(2).

4 Empirical study: Industrial Production Index of Belgium

We illustrate our proposed methodology by studying the monthly IPI of Belgium[3] from January 1960 to March 2023 (759 observations). We consider the Gaussian, mixture of normals and Student's t versions of the model in (1)-(2), see Blasques et al. (2024) and Harvey and Luati (2014), respectively, for the model specifications. The Student's t distribution has ν degrees of freedom

[3]Deseasonalized index of Total Industry Excluding Construction for Belgium (code BELPROINDMISMEI). Data were retrieved from FRED, Federal Reserve Bank of St. Louis. The data were transformed by taking the natural logarithm and multiplying by 100.

and scale parameter σ_1^2. The mixture of normal distributions has $J = 2$ components with weights w_1 and w_2, variances σ_1^2 and σ_2^2 and means zero. For each model, we select the lags p and q based on the Bayesian Information Criterion (BIC). The first 24 observations (2 years) are burned to allow the filters to converge to their correct paths.

Parameter estimates and information criteria for each model are shown in Table 1. The Gaussian model has the largest AIC and BIC values, while the mixed normal model clearly has the lowest AIC and BIC values. Hence, the overall in-sample preference is for the mixed normal model. The short-run component ψ_t has cyclical dynamics for all three models: all AR polynomials of the short-run component have complex roots.

For the Gaussian linear model, the estimate of κ, which is the long-run multiplier of the model in ARIMA representation, is 0.440, implying that a shock has a long-run effect that is less than half of its magnitude. For the nonlinear models, it is more difficult to directly interpret κ, due to the nonlinearity of the score. Figure 1 shows the value of κs_t for different values of ε_t, which shows the long-run impact of shocks in ε_t. For large values of ε_t, the long-term effect is lower for the non-Gaussian models than for the Gaussian model, while for values of ε_t close to zero, the value of κs_t is slightly higher for the non-Gaussian models.

Table 1: Parameter estimates with their standard errors (s.e.) using Belgian IPI data

	Gaussian		Mixed Normal		Student's t	
	$\widehat{\theta}_T$	s.e.	$\widehat{\theta}_T$	s.e.	$\widehat{\theta}_T$	s.e.
ω	0.207	*0.037*	0.181	*0.032*	0.200	*0.032*
κ	0.440	*0.042*	1.816	*0.208*	0.597	*0.068*
β_1	1.804	*0.059*	1.807	*0.048*	1.817	*0.045*
β_2	-0.845	*0.061*	-0.843	*0.050*	-0.852	*0.047*
α_1	0.061	*0.019*	0.441	*0.093*	0.138	*0.030*
σ_1^2	5.071	*0.265*	36.376	*5.208*	3.396	*0.237*
σ_2^2			3.820	*0.196*		
w_1			0.025	*0.009*		
w_2			0.975	-		
ν					7.611	*1.387*
LLH	-1639.54		-1583.78		-1591.65	
AIC	3291.08		3183.56		3197.31	
BIC	3318.68		3220.36		3229.51	

MLEs of parameters and corresponding log-likelihood values of model in (1)-(2) for different distributions, together with their corresponding information criteria AIC and BIC. The mixed normal distribution uses $J = 2$ mixture components.

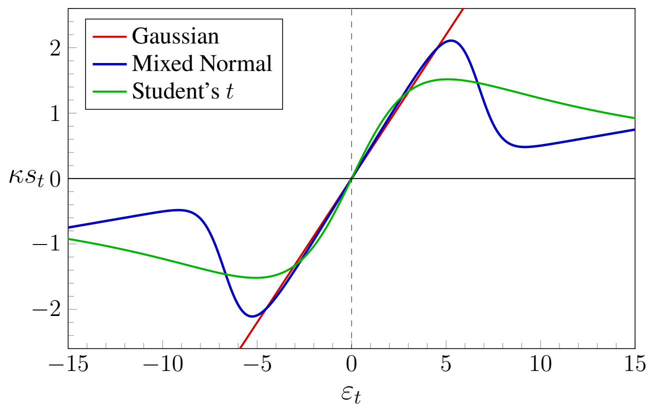

Figure 1: Increments of long-term component κs_t plotted as a function of ε_t for the three fitted models of Table 1.

The model-based decompositions of the observations into a permanent and transitory component are presented in Figure 2 together with recession periods in Belgium[4]. For the linear Gaussian model, we obtain exactly the original BN decomposition. The filtered trends of the Student's t and mixed normal models are smoother and less sensitive to outliers than the filtered trend of the linear Gaussian model. The robust BN decomposition delivered by the non-Gaussian score-driven models appear to be more satisfactory than the regular BN decomposition, because the long-term trend is traced more accurately. For all models, the cycle component shows a downward motion in virtually all of the recession periods.

[4]Indicated recession periods are the OECD based Peak through the Trough data for Belgium (code [BELREC]), retrieved from FRED, Federal Reserve Bank of St. Louis.

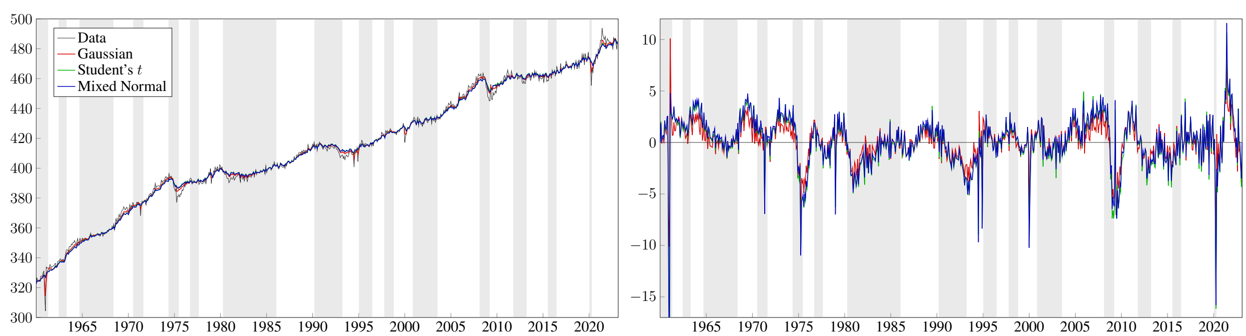

Figure 2: Filtered permanent component $\hat{\tau}_{t-1} - \omega$ (left panel) and transitory component $x_t - \hat{\tau}_{t-1} + \omega$ (right panel) corresponding to the estimates in Table 1. Shaded areas indicate Belgium's recession periods.

References

Anderson, H. M., Low, C. N., and Snyder, R. (2006). Single source of error state space approach to the Beveridge Nelson decomposition. *Economics Letters*, 91(1):104–109.

Beveridge, S. and Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle". *Journal of Monetary economics*, 7(2):151–174.

Blasques, F., van Brummelen, J., Gorgi, P., and Koopman, S. J. (2024). Maximum likelihood estimation for non-stationary location models with mixture of normal distributions. *Journal of Econometrics*, 238(1):105575.

Caivano, M. and Harvey, A. C. (2014). Time-series models with an EGB2 conditional distribution. *Journal of Time Series Analysis*, 35(6):558–571.

Clark, P. K. (1987). The cyclical component of U.S. economic activity. *Quarterly Journal of Economics*, 102:797–814.

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

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

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*. Cambridge University Press.

Harvey, A. C. and Jaeger, A. (1993). Detrending, stylised facts and the business cycle. *Journal of Applied Econometrics*, 8:231–247.

Harvey, A. C. and Luati, A. (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, 109(507):1112–1122.

Morley, J. C. (2011). The two interpretations of the Beveridge–Nelson decomposition. *Macroeconomic Dynamics*, 15(3):419–439.

Oh, K. H., Zivot, E., and Creal, D. (2008). The relationship between the Beveridge–Nelson decomposition and other permanent–transitory decompositions that are popular in economics. *Journal of Econometrics*, 146(2):207–219.

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics*, 34(5):2449–2495.

Watson, M. W. (1986). Univariate detrending methods and stochastic trends. *Journal of Monetary Economics*, 18:49–75.

Online Appendices

A BN decomposition of ARIMA model

Consider the ARIMA$(p, 1, q)$ model for $\{x_t\} \sim I(1)$

$$\phi_p(L)\Delta x_t = c + \theta_q(L)\varepsilon_t, \tag{A.1}$$

for $t = 1, \ldots, T$, where $\phi_p(L)$ and $\theta_q(L)$ are invertible polynomials of orders $p, q \in \mathbb{N}^0$ in the lag operator L, and $\Delta = 1 - L$. The disturbance ε_t is normally, independent and identically distributed with mean zero and variance σ^2. Assume that $\phi_p(L)$ is invertible, such that Δx_t has the Wold representation:

$$\Delta x_t = d + \vartheta_\infty(L)\varepsilon_t, \qquad \vartheta_\infty(L) = \phi_p(L)^{-1}\theta_q(L) = \sum_{k=0}^{\infty} \vartheta_k L^k,$$

where $d = \phi_p(1)^{-1}\, c$ and $\vartheta_0 = 1$. The BN decomposition relies on the lag polynomial identity given by $\vartheta_\infty(L) = \vartheta_\infty(1) + (1-L)\vartheta^\dagger_\infty(L)$, where the coefficients of $\vartheta^\dagger_\infty(L) = \vartheta^\dagger_0 + \vartheta^\dagger_1 L + \vartheta^\dagger_2 L^2 + \ldots$ are implicitly defined and given by $\vartheta^\dagger_j = -\sum_{k=j+1}^{\infty} \vartheta_k$, for $j = 0, 1, 2, \ldots$. The BN decomposition for the non-stationary time series x_t is then established as

$$x_t = a + d\,t + (1 - L)^{-1}\vartheta_\infty(1)\varepsilon_t + \vartheta^\dagger_\infty(L)\varepsilon_t,$$

where $a = x_0$ and $(1 - L)^{-1} = 1 + \sum_{k=1}^{\infty} L^k$. Equivalently, we can write $x_t = \tau^*_t + \psi^*_t$, where the trend τ^*_t and cycle ψ^*_t are specified as the dynamic processes

$$\tau^*_t = d + \tau^*_{t-1} + \vartheta_\infty(1)\varepsilon_t\,, \qquad \psi^*_t = \vartheta^\dagger_\infty(L)\varepsilon_t\,,$$

with $\tau^*_0 = a$.

B Calculation of the score of Gaussian location model

In this section we calculate the scaled score s_t of the Gaussian location model defined in (1)-(2). This is the scaled score function of the predictive likelihood contribution at time t with respect to time-varying parameters. In this case, it suffices to take the derivative with respect to the location μ_t, that is

$$s_t = S_t \cdot \partial \ell(y_t; \varphi) \,/\, \partial \mu_t, \qquad \ell(y_t; \varphi) = \log p(y_t | y_1, \ldots, y_{t-1}; \varphi),$$

where $p(\cdot)$ is the predictive probability density function, φ is the fixed unknown parameter vector and S_t is some scaling function. For now, we take $p(\cdot)$ to be the normal density function and $\varphi = (\omega\,,\, \kappa,\, \alpha_1,\, \ldots,\, \alpha_q,\, \beta_1,\, \ldots,\, \beta_p\,,\, \sigma^2)'$. When the score is scaled with the inverse of the Fisher information matrix. A regular choice of the scaling function is the inverse of the Fisher information, or a function thereof. We then have

$$\log p(y_t | y_1, \ldots, y_{t-1}; \varphi) = \text{constant} - \frac{1}{2}(y_t - \mu_t)^2 \,/\, \sigma^2, \qquad \partial \ell(y_t; \varphi) \,/\, \partial \mu_t = (y_t - \mu_t) \,/\, \sigma^2,$$

and

$$S_t = -\left[\, \mathrm{E}(\partial^2 \ell(y_t; \varphi) \,/\, \partial \mu_t^2) \,\right]^{-1} = \sigma^2.$$

Hence, the scaled score s_t is equal to $s_t \equiv y_t - \mu_t \equiv \varepsilon_t$.

C Equivalence ARMA and Gaussian score-driven model

C.1 Stationary ARMA case

The ARMA(p, q) model refers to an autoregressive moving average model of orders p and q, for any $p, q \in \mathbb{N}^0$, and, for a stationary univariate time series y_t, is represented by the dynamic process

$$\phi_p(L) y_t = c + \theta_q(L) \varepsilon_t, \tag{C.1}$$

for $t = 1, \ldots, T$, where the autoregressive and moving average polynomials are defined as $\phi_p(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ and $\theta_q(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$, with fixed and unknown coefficients $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$, and the disturbance ε_t is normally, identical and independently distributed with mean zero and variance σ^2, that is $\varepsilon_t \sim \text{NID}(0, \sigma^2)$. We assume that both lag polynomials $\phi_p(L)$ and $\theta_q(L)$ have their characteristic roots lying outside the unit circle. Hence, the process (C.1) for y_t is stationary.

The linear Gaussian ARMA(p, q) model in (C.1) can alternatively be represented as a stationary score-driven location model. For example the ARMA$(1, 1)$ model can alternatively be represented by the two equations

$$y_t = \mu_t + \varepsilon_t, \qquad \mu_{t+1} = \omega + \beta\mu_t + \alpha s_t, \tag{C.2}$$

where μ_t represents time-varying location, the normally distributed disturbance ε_t is as in (C.1), coefficients ω, α and β are treated as fixed unknown parameters, and s_t is the score, which we saw is equal to $y_t - \mu_t \equiv \varepsilon_t$ under Gaussian innovations. In case s_t is a stationary process, the stationary condition for the dynamic process of μ_t is simply $|\beta| < 1$.

It then follows that the score-driven location model (C.2) with a normally distributed disturbance $\varepsilon_t \sim \text{NID}(0, \sigma^2)$, reduces to the stationary ARMA$(1, 1)$ model. This equivalence becomes evident from the following steps

$$\begin{aligned} y_{t+1} &= \mu_{t+1} + \varepsilon_{t+1} \\ &= \omega + \beta\mu_t + \alpha s_t + \varepsilon_{t+1} \\ &= \omega + \beta y_t - \beta\varepsilon_t + \alpha s_t + \varepsilon_{t+1} \\ &= \omega + \beta y_t + (\alpha - \beta)\varepsilon_t + \varepsilon_{t+1}\,, \end{aligned}$$

which is the stationary ARMA$(1, 1)$ in model equations (C.1) with $p = q = 1$ and

$$c \equiv \omega, \qquad \phi_1 \equiv \beta, \qquad \theta_1 \equiv \alpha - \beta.$$

We notice that in case $\alpha \equiv \beta$, the specification reduces to an AR(1) process. These arguments extend naturally and straighforwardly to any combination of $p, q \in \mathbb{N}^0$ for the ARMA(p, q) model.

C.2 Nonstationary ARIMA case

Consider the Gaussian version of the model in (1)-(2), where we assume that the the roots of the characteristic equation of the AR polynomial are outside the unit circle. We will now demonstrate that this model is equivalent to an ARIMA$(p, 1, \max(p, q) + 1)$ model:

$$\begin{aligned}
& x_t = \tau_t + \psi_t + \varepsilon_t \\
\iff & (1-L)x_t = (1-L)\tau_t + (1-L)\psi_t + (1-L)\varepsilon_t \\
\iff & \Delta x_t = \omega + \kappa\varepsilon_{t-1} + (1-L)(1-\beta_1 L - \ldots - \beta_p L^p)^{-1}(\alpha_1 L + \ldots + \alpha_q L^q)\varepsilon_t + \Delta\varepsilon_t \\
\iff & \beta_p(L)\Delta x_t = \beta_p(1)\omega + \kappa\beta_p(L)\varepsilon_{t-1} + (\alpha_1 L + \ldots + \alpha_q L^q)\Delta\varepsilon_t + \beta_p(L)\Delta\varepsilon_t\,,
\end{aligned}$$

where $\beta_p(L) = 1 - \beta_1 L - \ldots - \beta_p L^p$. The final expression clearly shows that $\{x_t\}_{t\in\mathbb{N}}$ is an ARIMA$(p, 1, \max(p, q) + 1)$ process, see (A.1) for the definition of an ARIMA process. The drift parameter c is equal to $\beta_p(1)\omega$ and the autoregressive coefficients ϕ_i are equal to β_i for $i = 1, \ldots, p$. The moving average coefficients θ_i are functions of the parameters of the score-driven model. For example, the first-order MA coefficient is $\alpha_1 - \beta_1 + \kappa - 1$. In case $q < p$, the MA coefficients will be restricted, as there are only $q + 1$ 'free parameters' for the MA coefficients (κ and $\alpha_1, \ldots, \alpha_q$) and $p + 1$ MA lags. On the other hand, in case $q \geq p$ it can be shown straightforwardly that any combination of $q + 1$ MA coefficients can be obtained. It then also follows immediately that if $q \geq p$, the score-driven model can be made equivalent to an ARIMA$(p, 1, q^*)$ model with $q^* < p+1$ under particular restrictions on the parameters κ and $\alpha_1 \ \ldots, \alpha_q$, which ensure that certain MA lags cancel out. For example for $p = 1$ and $q = 1$, setting $\kappa = 1/(1 - \beta_1)$ and $\alpha_1 = -\beta_1^2/(1 - \beta_1)$ leads to an ARIMA$(1, 1, 0)$ process with autoregressive coefficient β_1.