Smooth transitions in volatility and correlation modelling

Annastiina Silvennoinen
NCER
School of Economics and Finance
Queensland University of Technology

Cast Seminar, Tampere University, 6 September, 2019

Based on joint adventures with

Timo Teräsvirta
CREATES
Department of Economics and Business
Aarhus University
C.A.S.E.
Humboldt-Universität zu Berlin

Cristina Amado
University of Minho, Braga
Outline

1. Smooth transition
2. Volatility
3. Correlation
4. Asymptotic theory and estimation
5. Specification tests
6. References
Smooth transition

Smooth transition is a convenient and flexible way of modelling a situation where a nonlinear shift in some part of the process is encountered.

For example, a smooth transition regression (STR) model introduces a transition in a linear regression model as

$$y_t = \phi' x_t + \psi' x_t G(\gamma, c; s_t) + e_t$$

where $x_t = (1, x_{1t}, \ldots, x_{pt})'$ is a vector of regressors and $\phi$ and $\psi$ are $p + 1$ parameter vectors.

The transition function commonly used is the logistic function

$$G(\gamma, c; s_t) = \left(1 + \exp\left\{-\gamma \prod_{i=1}^{k} (s_t - c_i)\right\}\right)^{-1}$$

where $\gamma > 0$ and $c_1 \leq c_2 \leq \ldots \leq c_k$ are identifying restrictions.

(See Teräsvirta (1998) and references therein.)
Smooth transition

Logistic function

- The transition variable $s_t$ either a weakly stationary stochastic variable or a deterministic function of time, $t/T$.
- The slope parameter $\gamma$ controls the speed of the transition. For example, when $\gamma \to \infty$, the transition becomes a step function. When $\gamma = 0$, the transition function $G$ becomes constant, and the model reduces to a linear specification.
- The transition speed and location(s) are estimated from the data.

![Logistic function graph]
Worth a mention

- Allows for a continuum of states between the two extremes.
- The transition occurs as a response to an observable indicator variable, or deterministically over time, and therefore has natural interpretation.
- Other transition functions are possible, as long as the transition function is bounded, continuous, and at least twice differentiable with respect to its parameters everywhere in the sample space. The choice may be motivated by the specific application and ease of interpretation.
Identification issue

Estimation of smooth transition models requires care.

If the data has been generated by a linear process, some of the model parameters ($\gamma$, $\psi$, and $c$ in the STR example) are not identified. For this reason, it is important to first test whether a nonlinear model is required. The test is set up as $H_0 : \gamma = 0$ against $H_1 : \gamma > 0$. However, the identification problem is not yet avoided, because the model is only identified under the alternative. As a consequence, the parameters cannot be consistently estimated and the standard asymptotic distribution theory for the classical test statistics does not work.
Smooth transition
Identification issue avoided

To circumvent the identification issue, a standard way forward is to follow Luukkonen, Saikkonen, and Teräsvirta (1988). They expand the logistic transition function as Taylor series around the point $\gamma = 0$.

For example, in the STR example and using first-order expansion, the model becomes

$$y_t = \beta'_0 x_t + \beta'_1 (x_t s_t) + e^*_t$$

where the $p + 1$ parameter vector $\beta_1$ is of the form $\gamma \tilde{\beta}_1$ and the error term $e^*_t = e_t + R\psi'x_t$ where $R$ is the approximation remainder which is zero under the null of $\gamma = 0$.

Therefore, the original null $H_0 : \gamma = 0$ can now be written as $H'_0 : \beta_1 = 0$. A standard LM test can now be applied, requiring estimation of the linear model only.

(See Teräsvirta (1998) and references therein for more details.)
In general, the modelling cycle should follow the sequence:

- **Specification**
  - initial test of linearity / parameter constancy
- **Estimation**
- **Evaluation**
  - residual diagnostics
  - detection of further nonlinearities → Specification
Eye-balling WIG20 variance: time-varying, clustering, memory, fat-tailed
Volatility
Modelling heteroskedasticity

Separate the time-varying variance $h_t$ from the error term:

$$\varepsilon_t = h_t^{1/2} z_t$$

where $z_t \sim iid(0, 1)$.

Autoregressive Conditional Heteroskedasticity (ARCH) was born (Engle, 1982)

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-1}^2$$

An upgrade, Generalised Autoregressive Conditional Heteroskedasticity (GARCH), was released (Bollerslev, 1986)

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-1}^2 + \sum_{i=1}^{p} \beta_i h_{t-1}$$

Asymmetric effects captured in GJR–GARCH (Glosten, Jagannathan, and Runkle, 1993)

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{i,t-1}^2 + \sum_{i=1}^{q} \kappa_i I(\varepsilon_{t-1} < 0) \varepsilon_{i,t-1}^2 + \sum_{i=1}^{p} \beta_i h_{t-1}$$
\( h_t = 0.05 + 0.03 \varepsilon_{t-1}^2 + 0.95 h_{t-1}, \quad \varepsilon_t = h^{1/2} z_t, \quad z_t \sim iidN(0,1) \)
Autocorrelation of absolute returns

WIG20

Too much memory for a standard GARCH to handle. Long memory or structural changes?
Multiplicative variance decomposition (MTV–GARCH)

Amado and Teräsvirta (2013): TVV–GARCH

\[ \varepsilon_{it} = g_{it}^{1/2} h_{it}^{1/2} z_{it} \]

where \( z_{it} \sim iid(0, 1) \), \( i = 1, \ldots, N \).

\[ g_{it} = g_{i}(t/T) = \delta_{i0} + \sum_{j=1}^{r_i} \delta_{ij} G_{ij}(\gamma_{ij}, c_{ij}; t/T) \quad i = 1, \ldots, N \]

where \( \delta_{i0} > 0 \) is a known constant, and \( c_{ij} = (c_{ij1}, \ldots, c_{ijK_{ij}}) \). \( G_{ij} \) is the generalised logistic function

\[ G_{ij}(\gamma_{ij}, c_{ij}; t/T) = \left( 1 + \exp \left\{ -\gamma_{ij} \prod_{k=1}^{K_{ij}} (t/T - c_{ijk}) \right\} \right)^{-1} \]

where \( \gamma_{ij} > 0 \) and \( c_{ij1} \leq \ldots \leq c_{ijK_{ij}} \) are required for identification.

\[ h_{it} = \alpha_{i0} + \alpha_i w_{i,t-1}^2 + \kappa_i I(w_{i,t-1} < 0) w_{i,t-1}^2 + \beta_i h_{i,t-1} \quad i = 1, \ldots, N \]

where \( w_{it} = \varepsilon_{it}/g_{it}^{1/2} \), and the usual constraints apply.
Volatility

Estimated slow-moving variance

Silvennoinen (QUT)

Smooth transition modelling

Tampere University 2019
Estimated GARCH component
Autocorrelation of standardised absolute returns
Volatility

MTV vs Spline

Silvennoinen (QUT)
Joint modelling of individual volatility dynamics and interaction between series. For surveys, see Silvennoinen and Teräsvirta (2009) and Amado, Silvennoinen, and Teräsvirta (forthcoming).

Rough breakdown of main modelling paths:

- Conditional covariances
  - Constant (CCC), Bollerslev (1990)
  - Markov-Switching (MSCC), Pelletier (2006)
  - Dynamic Equicorrelation (DECO), Engle and Kelly (2012)

- Conditional correlations
The observable stochastic $N \times 1$ vector $\varepsilon_t$ is decomposed as is customary

$$\varepsilon_t = H_t^{1/2}z_t = S_tD_tz_t, \quad z_t \sim \text{indep}(0, P_t)$$

where $z_t$ are independent with $E(z_t) = 0$ and $Cov(z_t) = P_t$. It follows that $\zeta_t = P_t^{-1/2}z_t \sim iid(0, I)$.

Further, $S_t = \text{diag}(g_{1t}^{1/2}, \ldots, g_{Nt}^{1/2})$, and $D_t = \text{diag}(h_{1t}^{1/2}, \ldots, h_{Nt}^{1/2})$ contains the conditional standard deviations of $w_t = S_t^{-1}\varepsilon_t = \text{diag}(\varepsilon_{1t}/g_{1t}^{1/2} \ldots \varepsilon_{Nt}/g_{Nt}^{1/2})$.

In other words, the conditional covariance matrix of $\varepsilon_t$ can be written as

$$H_t = S_tD_tP_tD_tS_t$$

where $P_t$ is a positive definite, valid correlation matrix for all $t$. 
Smooth transition (STCC) or time-varying (TVCC) conditional correlations

$P_t$ is a convex combination of two positive definite correlation matrices, $P(1)$ and $P(2)$.

Silvennoinen and Teräsvirta (2009, 2015):

$$P_t = (1 - G(\gamma, c; s_t))P(1) + G(\gamma, c; s_t)P(2)$$

where $P(1)$ and $P(2)$ are two positive definite correlation matrices, and $c = (c_1, \ldots, c_K)$. $G$ is the generalised logistic function

$$G(\gamma, c; s_t) = \left(1 + \exp \left\{ -\gamma \prod_{k=1}^{K} (s_t - c_k) \right\} \right)^{-1}$$

where $\gamma > 0$ and $c_1 \leq \ldots \leq c_K$ are required for identification. The transitioning takes place according to the values of the transition variable $s_t$, which may also be a deterministic time trend $t/T$. 
Australian banks: The Big-4

ANZ

CBA

NAB

WBC
TVV Estimates

ANZ

CBA

NAB

WBC
STCC Estimates

\[ \gamma = 12.131 \]
\[ c = 0.6443 \]

Transition start: end of 2000
Transition end: end of 2010

<table>
<thead>
<tr>
<th></th>
<th>ANZ</th>
<th>CBA</th>
<th>NAB</th>
<th>WBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{(1)} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANZ</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBA</td>
<td>0.4974</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAB</td>
<td>0.5158</td>
<td>0.5369</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>WBC</td>
<td>0.6062</td>
<td>0.5081</td>
<td>0.4807</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ANZ</th>
<th>CBA</th>
<th>NAB</th>
<th>WBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{(2)} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANZ</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBA</td>
<td>0.8308</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAB</td>
<td>0.8551</td>
<td>0.8254</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>WBC</td>
<td>0.8663</td>
<td>0.8607</td>
<td>0.8614</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Asymptotic theory

The theory is essential in model specification and all inference in general. The most general asymptotic results are limited to ARMA-ECCC models (Ling and McAleer, 2003).

Engle and Kelly (2012): “... a rigorous analysis of asymptotic theory for multivariate GARCH processes remains an important unanswered question.”
It has been shown that the TVV–model maximum likelihood estimates are consistent and asymptotically normal (Amado and Teräsvirta, 2013), and that the TVV–TVCC–model estimates are consistent and asymptotically normal (Silvennoinen and Teräsvirta, 2019). Then, assuming the GARCH process is weakly stationary and the regularity conditions hold, and following Song, Fan, and Kalbfleisch (2005), the maximum likelihood estimates of the GARCH parameters are also consistent and asymptotically normal.

Estimation of all parameters simultaneously is numerically demanding. Maximisation by parts is therefore recommended. Iterative cycle between the TVV-parameters, correlation parameters, and GARCH parameters is continued until convergence. The final maximum likelihood estimator is consistent and asymptotically normal (Song et al., 2005).
The MTV–STCC–GARCH model is rather general and nests many, more specific submodels. Fitting a more general model when the data is generated by a smaller submodel will lead to inconsistent parameter estimates, due to unidentified parameters. Systematic model building approach, starting from a small model and using statistical inference to specify nonlinearities before estimation, is therefore recommended. Misspecification tests are then used to evaluate the model, to assess whether the end product is acceptable, or if additional nonlinearities are detected.

The tests are LM-type tests, only requiring the null model to be estimated. Due to the unidentified parameters in the alternative, the nonlinear transition function $G_t$ is approximated by a Taylor expansion. The order of expansion is typically 1, 2, or 3.
There are plenty of specification and evaluation tests available for the univariate MTV–GARCH component: Amado and Teräsvirta (2013, 2017), Silvennoinen and Teräsvirta (2016), Hall, Silvennoinen, and Teräsvirta (2019); Amado, Silvennoinen, and Teräsvirta, (2017), Lundbergh and Teräsvirta, (2002), and others.

In the correlation component, it has been mostly quiet, even when it comes to testing whether correlations are constant or time-varying. The tests of Silvennoinen and Teräsvirta (2009, 2015) approximate many forms of nonlinear time-varying correlation structures, and for that reason are powerful, however, only up to moderate dimensions. The neural networks based test of Péguin-Feissolle and Sanhaji (2016) as well as their Taylor expansion based test perform similarly.
STCC-based test

The original test of constant correlations (and its extensions to further specifications) is based on the correlation specification

$$P_t = (1 - G(\gamma, c; s_t))P_{(1)} + G(\gamma, c; s_t)P_{(2)}$$

where the logistic function $G$ is expanded into Taylor series around the null hypothesis of $\gamma = 0$, giving

$$P^* = P_{(0)} + s_t P_{(1)} + s_t^2 P_{(2)} + P_{(R)}$$

where $P_{(R)}$ is a residual matrix (which is null under $H_o$), and $P^*_{(1)}$ and $P^*_{(2)}$ are symmetric and have zeros on their diagonal.

The null hypothesis becomes $H'_o : P^*_{(1)} = P^*_{(2)} = 0$, which has dimension $N(N - 1)$. 
Parsimonious test

Assume the correlation matrix $P_t$ is rotation invariant, and break it into parts using the eigenvalue decomposition:

$$P_t = Q \Lambda_t Q'$$

where $Q$ holds the eigenvectors, $QQ' = I$,

$$\Lambda_t = (1 - G(\gamma, c; s_t))(\Lambda - \Lambda^*) + G(\gamma, c; s_t)(\Lambda + \Lambda^*)$$

and $G$ is the generalised logistic function

$$G(\gamma, c; s_t) = \left(1 + \exp \left\{ -\gamma \prod_{k=1}^{K} (s_t - c_k) \right\} \right)^{-1}$$

When $\gamma = 0$, the correlation matrix is constant $P = Q\Lambda Q'$, and $\Lambda$ holds the eigenvalues of $P$, $\lambda_i, i = 1, \ldots, N$, $\lambda_i > 0, \forall i$, and $\sum \lambda_i = N$. 
Developing the generalised logistic function $G$ into a Taylor series around the null hypothesis of $\gamma = 0$ yields

$$\Psi_t = \Psi(0) + s_t \Psi(1) + s_t^2 \Psi(2) + \Psi(R)$$

where $\Psi(R)$ is a residual matrix (which is null under $H_0$), $\Psi(1)$ and $\Psi(2)$ are diagonal parameter matrices, and $\Psi(0) = \Lambda$ when $H_o$ holds.

The diagonal elements of $\Psi(j)$ are of the form $\psi(j,i) = \gamma \tilde{\psi}(j,i)$, $j = 1, 2$, $i = 1, \ldots, N$ with $\tilde{\psi}(j,i) \neq 0$. Hence, the new $H'_o$ becomes $\Psi(1) = \Psi(2) = 0$.

The dimension of the proposed parsimonious test is $2N$. 
If the test rejects the null, then what?

Because the test is constructed on an approximation of the alternative, and various other nonlinear alternatives could reasonably have the same or very similar linear approximation, the rejection of the null is not interpreted as definite support for the particular model for \( \Lambda_t \) we used.

In fact, the matrix \( P_t \) (as defined previously in the parsimonious test) is not generally a valid correlation matrix when \( P_t \neq P \).

However, the model may be estimated (for instance, by requiring correlation targeting by fixing the eigenvectors in \( Q \) and eigenvalues in \( \Lambda \) to match the unconditional correlation matrix, and placing suitable restrictions on the parameters in \( \Lambda^* \) to ensure positive definiteness of \( P_t \)) and the implied correlations obtained by the transformation

\[
C_t = (P_t \odot I)^{-1/2} P_t (P_t \odot I)^{-1/2}
\]
Smooth transition equicorrelation (STEC)

Returning to the spectral decomposition

\[ P_t = Q \Lambda_t Q' \]

where

\[ \Lambda_t = (1 - G(\gamma, c; s_t))(\Lambda - \Lambda^*) + G(\gamma, c; s_t)(\Lambda + \Lambda^*) \]

One case when this specification yields a valid correlation matrix at each point in time is when the correlations are equal.

The eigenvectors and eigenvalues are analytically tractable, and matrix inversions are avoided. Specifically,

\[ \Lambda_t = \text{diag}(1 + (N - 1)\rho_t, 1 - \rho_t, \ldots, 1 - \rho_t) \]

where

\[ \rho_t = (1 - G(\gamma, c; s_t))(\rho - \rho^*) + G(\gamma, c; s_t)(\rho + \rho^*) \]

and restrictions \(|\rho^*| < (N - 1)^{-1} + \rho\) and \(|\rho^*| < 1 - \rho\) must hold.
In the STEC model, setting $\gamma = 0$ will reduce the model to a constant equicorrelation model, with

$$\Lambda = \text{diag}(1 + (N - 1)\rho, 1 - \rho, \ldots, 1 - \rho)$$

Following the same approach as before, the Taylor expansion yields an approximation for the eigenvalues:

$$\Psi_t = \text{diag}(1 + (N - 1)\psi_t, 1 - \psi_t, \ldots, 1 - \psi_t)$$

where

$$\psi_t = \psi(0) + \psi(1)s_t + \psi(2)s_t^2 + \psi(R)$$

and the new $H'_o$ is $\psi(1) = \psi(2) = 0$. Under $H_o$, $\Psi_t = \Lambda$. 
Properties of the tests

The smooth transition based tests rely on having information of the true transition variable. In practice, this is rarely the case, and therefore the tests should be viewed as conditional of a particular candidate for a transition variable.

For this reason, the tests typically have high power against models where correlations are driven by the same variable as the one used in the test, and close to no power when this is not the case. The advantage of this feature of the tests is that they allow for testing with a multitude of candidates for transition variables, and deliver information whether those candidates are indicators for time-variation in correlations.

The STCC based test begins to lose power as the dimension of the model increases, while the parsimonious test retains its power for longer. Both the parsimonious and the STEC based test work well in high dimensions and non-equicorrelated models, as long as the strength and direction of change in the correlations are similar.


Selected References II


