
Lenard Lieb¹ and Bertrand Candelon²

- Draft Version -

Abstract

This paper develops an estimation and testing framework for a threshold VECM, where short-run dynamics are regime dependent and are driven by an exogenous, stationary and ergodic threshold variable. We modify the traditional Wald-type test for linearity and derive its asymptotic distribution, which turns out to be non-standard, but similar to the one presented in Andrews (1993). We apply this approach to investigate the stability of money demand in the U.S. In order to account for the diversity of the potential determinants of money holdings, a broad scope of money demand systems is analyzed. The implementation of the linearity test allows us to confirm that money demand is not only suffering from structural breaks, but is driven by regimes governed by the stance of the economy.

Keywords: Non-linearity; Co-Integration; Threshold Vector Autoregression; Money Demand.
1 Introduction

Following Tong (1990) and Tsay (1998), several papers stressed that macroeconomic dynamics (univariate or multivariate) may be non-linear and regime-dependent. To integrate such a feature, the traditional class of VAR (or VECM) models appear to be inadequate, as they are by essence linear, hence unable to offer regime-specific dynamics. Initial tries to break down linearity has been provided by the Markov-switching literature (see Hamilton, 1989). This model framework provides estimates of state dependent autoregressive coefficients and/or variance, as well as the transition matrix, which provides the probability of changing from one regime to another. However, multivariate Markov-switching models have several drawbacks, the most serious one being that they are difficult to interpret, as the threshold variable and the cut-offs are not observed. In contrast, threshold VARs, or so called threshold co-integrated models (T-VECM), do not suffer from such an identification problem, as the variable governing the non-linearity is directly imposed. The latter class of models allows to control for threshold-effects driving the error correction term (e.g. Hansen and Seo, 2002 or Seo, 2008). Hence, they cover non-linearities in the long-run dynamics, while those due to short-run modifications are not appropriately addressed.\(^3\) This appears to be inadequate for modeling many macroeconomic dynamic relations. For instance, using a flexible least square method approach, Lütkepohl (1993) shows that money demand instability should not be associated with long-run changes but should be instead related to modifications in the short-run dynamics. Moreover, complementary theoretical models such as DSGEs always consider variables in deviation with respect to a fixed (therefor linear) steady state, when analyzing the effect of a shock.

As the econometric literature has been mainly focused on threshold co-integration, only a few papers stressed the importance of non-linearities in the equilibrium adjustment and/or the short-

\(^3\)The literature provides two important extensions of "standard" T-VECMs. Gonzalo and Pitarikis (2007) present a model where threshold-effects in the error correction term are allowed to be caused by external variables. Seo (2011) derives a framework which allows regime-specific variations in the short-run dynamics, however, regimes are also determined by the error correction term.
run dynamics in the context of macroeconomic modeling. Recent examples include Lütkepohl et al. (1999) or Teräsvirta and Eliasson (2001), both based on smooth transition regression (STR). In contrast to threshold models, STR models are not piecewise linear, since they provide an estimate of a continuous (in the transition variable) transition function. Hence, they might be more adequate to model non-linearities in the data, and thus the true DGP. In some situations, however, piecewise linearity remains still an attractive model feature. Sometimes researchers are rather interested in identifying exact regimes (e.g. business cycle dating) or in studying regime dependent dynamics.

Kurita and Nielsen (2009), for instance, present a model that allows for structural breaks in the short-run parameters. This approach, however, might not be adequate for investigating any dynamic macroeconomic relationship, since regimes can often be recurrent.\footnote{For example, the episode of missing money in the late 1970’s is quite similar to the one that occurred in the early 2000’s.} Neglecting such a fact would lead to add a redundant structural break, penalizing hence the efficiency of the estimator. In contrast, the paper by Bec and Rahbek (2004) is undoubtedly closer related to our approach as they consider threshold effects that exhibit short-run dynamics. However, they assume a threshold governed by a Markov process, leading to the drawbacks mentioned earlier.

Motivated by the fact that non-linearities in economic data are associated with short-run dynamics rather than long-run ones, we propose in this paper an estimation and testing strategy for a non-linear VECM, where the non-linearities are driven by threshold-effects in the short-run dynamics. In particular, our approach differs from the related literature, insofar as we do not assume any specific co-integration properties of the system. Moreover, we do not presume that non-linearities are triggered by the error correction term nor by a Markov process, but we consider a general external threshold variable, which could be any economic or financial variable that is stationary and ergodic. This allows for a straightforward interpretation of both: the identified regimes and the threshold. Within this framework, we elaborate a Wald-type test for linearity and derive its asymptotic properties.
As an empirical application, we use our model framework to analyze the evolution of the demand for broad money in the U.S. As several empirical studies suggest,\(^5\) money demand instability is rather associated with short-run changes than with modifications in the long-run equilibrium behavior. We investigate the origin of this non-linearity by testing for threshold effects resulting from movements in the business cycle. Our findings confirm the existence non-linear effects in the short-run dynamics. The pattern of the identified regimes allows to draw the conclusion that money demand differs in times of recession compared to “normal” times.

The rest of the paper is organized as follows. Section 2 presents the non-linear VECM framework. In section 3 a quasi-ML estimation approach is derived, while the Wald-type test for linearity as well as its asymptotic properties are developed in section 4. Size and power characteristics of the test are studied in section 5. In section 6 we apply this framework to investigate whether money demand in the U.S. has been stable or not. This includes a battery of robustness checks. Section 7 concludes.

2 A Model with Short-Run Threshold Effects

The model framework used in the following can be built on a version of the non-linear VECM proposed by Bec and Rahbek (2004). As they assume that the threshold is triggered by a Markov chain their model is formulated as follows.

\[
\Delta y_t = s_t \left[ \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma^A_i \Delta y_{t-i} \right] + (1 - s_t) \left[ \alpha' \beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma^B_i \Delta y_{t-i} \right] + u_t
\]

(1)

where \(\{y_t\}\) is a \(K\)-dimensional time series, \(\beta\) captures the common long-run equilibrium dynamics, \(\alpha\) the adjustment to the long-run equilibrium, and \(\Gamma^A, B\) the regime-specific short-run dynamics. As usual, \(p\) is the lag order, \(u_t\) is the error term assumed to be a zero mean i.i.d. sequence of random vectors with covariance \(\Sigma_u > 0\), and the state variable \(s_t\) is governed by a Markov chain.

\(^5\)e.g. Lütkepohl et al. 1999; Tersvirta and Eliasson, 2001; Sarno et al., 2003 or Calza and Zaghini, 2009.
It is noticeable that in this representation, $\alpha$, which represents the adjustment to the long-run equilibrium is not regime dependent.

As mentioned earlier threshold-effects triggered by Markov-switching allow for a straightforward modeling of the transition from one regime to another, they involves, however, the problem of identification as the threshold is not governed by an interpretable variable, e.g. a growth rate. Contrary to the MS-VECM, we do not aim at estimating the probability of transition from one regime to another, but instead the threshold variable and the value of the cut-off above which a particular regime prevails.

As one of the central goals of this paper is to provide an estimator of an interpretable threshold and not to model the transition, we can elaborate from model (1) and write a short-run threshold VECM (SR-TVECM) as:

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i^A \Delta y_{t-i} I(z_{t-d} \leq \nu) + \sum_{i=1}^{p-1} \Gamma_i^B \Delta y_{t-i} I(z_{t-d} > \nu) + u_t, \tag{2}$$

where $\Pi := \alpha \beta'$ is a $K \times r$, $0 < r < K$ matrix governing the long-run equilibrium dynamics and its equilibrium adjustment. $I(\cdot)$ denotes the indicator function. Again, non-linearities exist due to regime-specific short-run dynamics. Here, the determination of the regime to which $\Delta y$ belongs to at time $t$, depends on a (external) threshold variable $z_t$ with delayed impact $d$ being less equal and greater, respectively, than the threshold $\nu$.

## 3 Model Estimation

Since the estimation and testing framework derived in the following is based on the SR-TVECM as formulated in (2), we introduce below some basic conditions we will need throughout the rest of the paper.

**Assumption 1.**

1. The model is of lag-order $p > 1$ and the process $\{y_t\}$ is integrated of order 1.
2. No deterministic terms are included.

3. The threshold variable \( z_t \) is strictly stationary and ergodic.

4. The threshold space \( \Upsilon = [\upsilon_L, \upsilon_U] \subset \mathbb{R} \) is a closed and bounded subset of the given sample sequence \( \{z_t\}_{t=1}^T \).

5. The errors \( u_t \) are assumed to be iid Gaussian, hence \( E|u_t|^4 < \infty \) holds for some \( r \geq 1 \) and for \( i = 1, ..., K \forall t \).

6. Moreover, it is assumed that the residuals are conditionally homoskedastic, i.e. \( E(u_t \upsilon_t'|\mathcal{F}_{t-1}) = \Sigma_u \), a.s., where \( \mathcal{F}_{t-1} \) denotes the sigma field generated by current and lagged values of \((y_t, u_t)\).

Under the assumption that the errors \( u_t \) are iid Gaussian, we propose in the following an estimation procedure based on the ML-principle. In order to elaborate on the estimation strategy, the model (2) is rewritten in matrix notation as

\[
\Delta Y = \Pi Y_{-1} + \Gamma^A \Delta X J^A(\upsilon) + \Gamma^B \Delta X J^B(\upsilon) + U, \tag{3}
\]

where \( \Delta Y := [\Delta y_1, ..., \Delta y_T], Y_{-1} := [y_0, ..., y_{T-1}], \Gamma^A,B := [\Gamma^A_1, ..., \Gamma^A_{p-1}], \Delta X := [\Delta X_0, ..., \Delta X_{T-1}] \) with \( \Delta X_{t-1} := [\Delta y'_{t-1}, ..., \Delta y'_{t-p+1}]' \), \( U := [u_1, ..., u_T] \), and \( J^A,B(\upsilon) \) are \((T \times T)\) diagonal matrices where the \((j-th, j-th)\) element is determined by \( I(z_{j-d} \leq \upsilon) \), and \( I(z_{j-d} > \upsilon) \), respectively.

As we assume a single and unique steady state, the reduced rank parameter matrix \( \Pi \) is not regime dependent and it is straightforward to concentrate (2) with respect to the short-run parameters by replacing \( \Pi \) by its LS estimator for any given \( \Gamma^A \) and \( \Gamma^B \),

\[
\Pi(\Gamma^A, \Gamma^B) = (\Delta Y - \Gamma^A \Delta X J^A(\upsilon) - \Gamma^B \Delta X J^B(\upsilon)) Y_{-1}(Y_{-1}'Y_{-1})^{-1}. \tag{4}
\]

Using an appropriate projection matrix, we can now write the concentrated form of (2) in matrix notation as

\[
\mathcal{Y} = \Gamma^A \mathcal{X} J^A(\upsilon) + \Gamma^B \mathcal{X} J^B(\upsilon) + U^* \tag{5}
\]
where \( \mathcal{Y} = [\mathbf{y}_1, ..., \mathbf{y}_T] := \Delta \mathbf{Y}(I_T - \hat{\mathbf{Y}}) \) with \( \hat{\mathbf{Y}} := Y_{-1}'(Y_{-1}'Y_{-1})^{-1}Y_{-1} \), and \( \mathbf{U}^* := [u^*_1, ..., u^*_T] \).

From a computational point of view it is convenient to hold the threshold \( \upsilon \) fixed, so that we can compute \( \hat{\Gamma}^A(\upsilon) \) and \( \hat{\Gamma}^B(\upsilon) \) as constrained LS estimators. This is just OLS regression yielding the parameter estimates

\[
\hat{\Gamma}^A(\upsilon) = \mathcal{Y}(\mathbf{X}^A(\upsilon)') (\mathbf{X}(\mathbf{X}^A(\upsilon)'))^{-1}, \quad (6)
\]

\[
\hat{\Gamma}^B(\upsilon) = \mathcal{Y}(\mathbf{X}^B(\upsilon)') (\mathbf{X}(\mathbf{X}^B(\upsilon)'))^{-1}. \quad (7)
\]

The threshold \( \upsilon \) in (5) can be estimated by maximum likelihood (ML), if \( u^*_t \sim \mathcal{N}(0, \Sigma_{u^*}) \). For this purpose the Gaussianity of the residuals \( u^*_t \) is established by the following

**Lemma 2.** Under Ass. 1, \( u_t \) and \( u^*_t \) are asymptotically equivalent.

**Proof.** see appendix. \[ \square \]

Hence it follows that \( u^*_t \sim \mathcal{N}(0, \Sigma_u) \) and that \( \hat{\Sigma}_{u^*} = \hat{\Sigma}_u = \hat{\mathbf{U}}^* \hat{\mathbf{U}}^*'/T \) is a consistent estimator of the variance-covariance matrix.

Based on the constrained parameter estimates \( \hat{\Gamma}^A(\upsilon), \hat{\Gamma}^B(\upsilon), \) and \( \hat{\Sigma}_{u^*}(\upsilon) := \hat{\mathbf{U}}^*(\upsilon)\hat{\mathbf{U}}^*(\upsilon)'/T, \) where \( \hat{\mathbf{U}}^*(\upsilon) = \mathcal{Y} - \hat{\Gamma}^A(\upsilon)\mathbf{X} - \hat{\Gamma}^B(\upsilon)\mathbf{X} \), the log-likelihood function for a sample of size \( T \) is given by

\[
\mathcal{L}_T(\upsilon) = -\frac{KT}{2} \log(2\pi) - \frac{T}{2} \log \left( \det \hat{\Sigma}_u(\upsilon) \right) - \frac{1}{2} \text{tr} \left[ \hat{\Sigma}_{u^*}(\upsilon)^{-1} \hat{\mathbf{U}}^*(\upsilon) \right] = -\frac{KT}{2} \log(2\pi) - \frac{T}{2} \log \left( \det \hat{\Sigma}_u(\upsilon) \right) - \frac{TK}{2} \quad (8)
\]

Hence, the ML-estimator of the threshold \( \hat{\upsilon} \) is found to be the minimizer of

\[
\hat{\upsilon} = \arg \min_{\upsilon \in \Upsilon} \det \hat{\Sigma}_u(\upsilon). \quad (9)
\]
Since this optimization task cannot be solved analytically, numerical approximations have to be performed. As we are considering only a one-dimensional threshold space, a grid search is a simple and fast solution.\footnote{Popular algorithms, like the method of steepest descent, cannot be used here, since the function to be maximized is not smooth.}

In a second step, we can estimate the reduced rank matrix \( \Pi = \alpha \beta' \) by EGLS based on the ML estimators \( \hat{\Pi}(\hat{\Gamma}^A(\hat{\upsilon}), \hat{\Gamma}^B(\hat{\upsilon})) \).\footnote{See Saikkonen (1992).} Note that the non-linearity in the short-run parameters does not affect the estimation of \( \Pi \), since we have, due to the indicator function, that \( \hat{\Gamma}\Delta X = \hat{\Gamma}^A \Delta X J^A(\upsilon) + \hat{\Gamma}^B \Delta X J^B(\upsilon) \), and hence \( \hat{\Pi}(\hat{\Gamma}) = \hat{\Pi}(\hat{\Gamma}^A(\upsilon), \hat{\Gamma}^B(\upsilon)) \), where \( \hat{\Gamma} \) is the parameter estimate resulting from the linear regression in model (5). Moreover, since \( \Pi \) has been replaced by the optimal matrix for any given \( \Gamma \) (or equivalently for any given matrices \( \Gamma^A(\upsilon) \) and \( \Gamma^B(\upsilon) \)) the same estimator would have been obtained if \( \Pi \) would have not be concentrated out first. Consequently, the standard procedures for determining the co-integration rank can be used.\footnote{Usually, also the delay parameter \( d \) is unknown and has to be determined somehow. Hansen (1997) suggests, within a simple TAR framework, to use the LS principle and estimate \( d \) together with the other parameters. In the proposed multivariate estimation-framework this task becomes quickly computationally expensive as the grid size increases. Instead, we extended standard model-selection criteria to determine both, the lag-order and the delay parameter. Kapetios (2001) has shown that this approach delivers consistent results, using standard model selection criteria (AIC,HQ,SC).}

**Remark 3** In this section we have described an algorithm how to compute the quasi-ML estimator of a two-regime SR-TVECM. It should be stressed that we did not derive asymptotic inference of the estimated threshold parameter. It remains a conjecture that \( \hat{\upsilon} \) converges to its true value \( \upsilon \) at rate \( T \) as it does in stationary models (see e.g. Gonzalo and Pitarakis, 2002). Although we do not provide a proof, it is reasonable to guess that the convergence properties of the threshold estimate are similar to those found for stationary models because the concentrated version of the model we use for estimation is stationary (see below) and \( u_t^* \overset{P}{\rightarrow} u_t \).
4 Testing for Short-Run Threshold Effects

4.1 The Test Statistic

Let $\mathcal{H}_0$ denote the class of standard linear VECMs and $\mathcal{H}_1$ the class of a two-regime threshold model as described in (2). As both model classes are nested, $\mathcal{H}_0$ contains all models in $\mathcal{H}_1$ which satisfy $\Gamma^A = \Gamma^B$.

We are interested in testing whether the true DGP follows a linear model $\mathcal{H}_0$, or if it is rather described by a non-linear model as under the $\mathcal{H}_1$. For this purpose we opt for a Wald-type test. The test statistic as stated below and its asymptotic distribution is additionally based on the following assumption about the threshold variable.

**Assumption 4**. $z_t$ has a continuous distribution function $F$. Since $F$ is monotonically increasing we can make use of the inequality $I(z_{t-d} \leq v) = I(F(z_{t-d}) \leq F(v))$. Moreover, the threshold space $\Upsilon = [\nu_L, \nu_U] \subset \mathbb{R}$ is closed and bounded, such that $F(\nu_L) > 0$ and $F(\nu_U) < 1$ (usually it is assumed that $F(\nu_U) = 1 - F(\nu_L)$). Hence, $F(\Upsilon)$ has closure in $(0, 1)$.

Since we are only interested in threshold-effects in the short-run dynamics, the testing procedure can be based on the model as formulated in (5). The following lemma will be needed to establish the Wald-test.

**Lemma 5**. Let the $K$-dimensional series $\{y_t\}$ generated by the error correction model in (2) be $I(1)$, then the series $\{w_t = (y_{t}, x_t)\}$ that satisfy the regression relationship in (5) are asymptotically stationary under the null if Ass. 1 holds.

**Proof.** see: Appendix

Based on the stationary regression in (5), we can now adopt a standard Wald-type test statistic for a sample of size $T$ and any given $v$. The statistic takes the following form

$$W_T(v) = T(\hat{\gamma}^A - \hat{\gamma}^B)' \left\{ \frac{\bar{X}^A(v)\bar{X}^A(v)'}{T} \right\}^{-1} \left\{ \frac{\bar{X}^B(v)\bar{X}^B(v)'}{T} \otimes \hat{\Sigma}_u^{-1} \right\} (\hat{\gamma}^A - \hat{\gamma}^B), \quad (10)$$
where $\hat{\gamma}^m := \text{vec}(\hat{\Gamma}^m)$, and $X^m(\nu) := X^m J^m(\nu), m = A, B$. The derivation of the functional form is left to the appendix.

If the true threshold parameter would be known (10) would be the test statistic, and if it is unknown it had to be replaced by its estimate. However, there exists no estimate of $\nu$ under the null and consequently we cannot use the statistic as it is defined in (10). A common method proposed in the literature to solve this “nuisance” parameter problem, is to use a weighting function $g(W_T(\nu))$ over all values in the threshold space. As a candidate for $g\{W_T(\nu): \nu \in \Upsilon\}$ Davies (1987) and Andrews (1993) suggest to take $\sup_{\nu \in \Upsilon} W_T(\nu)$ which corresponds to the tests derived from Roy’s type I principle (see: Roy, 1953). Andrews and Ploberger (1994) argue that an average (exponential) form of $W_T(\nu)$ might achieve higher power. Since the choice of the test statistic remains arbitrary due to the chosen weighting function we will use the $\sup W_T(\nu)$ (in the following the subscript is skipped sometimes for readability) as a benchmark to derive its asymptotic properties.

### 4.2 Asymptotic Distribution

In this subsection we provide the asymptotic distribution of the following test statistic under the null:

$$\sup W_T(\nu) = \sup_{\nu \in \Upsilon} \sqrt{T} (\hat{\gamma}^A - \hat{\gamma}^B)^T \left\{ \frac{X^B(\nu)X^B(\nu)'}{T} \left[ \frac{XX'}{T} \right]^{-1} \frac{X^A(\nu)X^A(\nu)'}{T} \otimes \hat{\Sigma}_u^{-1} \right\} \sqrt{T} (\hat{\gamma}^A - \hat{\gamma}^B).$$

Under Ass. 1 and 4 we can establish the following

**Proposition 6.** The asymptotic null distribution of the above statistic is similar to the one presented in Andrews (1993) and is given by

$$\sup_{\nu \in \Upsilon} W_T(\nu) \Rightarrow \sup_{s \in S} \Omega(s), \quad (11)$$

10
where \( S := [F(v_l), F(v_u)] \subset (0, 1) \) and \( \Omega(s) \) is a squared standardized tied-down Bessel process of order \( K^2(p - 1) \).

**Proof.** see appendix. ■

**Remark 7** The asymptotic distribution is free of nuisance parameters and does not depend upon moment functionals, hence critical values can be tabulated (see Andrews, 1993, p.840). However, since it is asymptotically pivotal it stands to reason that appropriate re-sampling procedures will lead to small-sample refinements. We explore the size and the power of the statistic in the next section.

## 5 Simulation Evidence

After having established the limiting behavior of the Wald-type statistic above, we will now investigate the adequacy of the asymptotic approximation stated in Prop. 6.

The Monte Carlo experiments are based on a bivariate non-stationary error-correction model which takes the following form under \( \mathcal{H}_0 \)

\[
\Delta y_t = \alpha \beta' y_{t-1} + \Gamma^A \Delta y_{t-1} + u_t, \tag{12}
\]

and

\[
\Delta y_t = \alpha \beta' y_{t-1} + \Gamma^A \Delta y_{t-1} I(z_{t-d} \leq v) + \Gamma^B \Delta y_{t-1} I(z_{t-d} > v) + u_t \tag{13}
\]

under the alternative. We fix \( \alpha = (-0.4, 0.1)' \), \( \beta' = (1, 2) \), and

\[
\Gamma^A = \begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.1 \end{bmatrix}, \quad \Gamma^B = \begin{bmatrix} 0.2 & -0.1 \\ 0.1 & 0 \end{bmatrix}.
\]

The threshold variable is chosen to follow an AR(1) process with uncorrelated standard normal innovations and persistence of 0.5 with delayed impact of one period. The trimming value is chosen such that \( F(v_l) = 0.15 \) and \( F(v_u) = 0.85 \), using a grid space of 0.01. Finally, the errors \( u_t \) are
assumed to follow a bivariate standard normal distribution. The sample size varies between 50, 100 and 250 observations. 500 residual bootstrap replications are computed for each replication and the results presented below are calculated from 1000 replications.

Next to exploring the power and size characteristics of the test statistic, another purpose of the simulation study is to investigate how accurately a finite sample distribution of the Wald statistic tracks their asymptotic counterpart. Table 1 compares critical values obtained by simulation based on 1000 bootstrap replications and the ones tabulated in Andrews (1993, p.840).

<table>
<thead>
<tr>
<th>T</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>15.80</td>
<td>17.85</td>
<td>22.77</td>
</tr>
<tr>
<td>100</td>
<td>14.89</td>
<td>16.88</td>
<td>21.08</td>
</tr>
<tr>
<td>250</td>
<td>13.99</td>
<td>15.70</td>
<td>20.03</td>
</tr>
<tr>
<td>500</td>
<td>14.37</td>
<td>16.13</td>
<td>20.35</td>
</tr>
<tr>
<td>Asympt. (Andrews)</td>
<td>14.31</td>
<td>16.45</td>
<td>20.71</td>
</tr>
</tbody>
</table>

Table 1: Critical values

We can clearly observe that, with increasing sample size, the critical values based on our DGP approach the asymptotic critical values very accurately.

Now, we explore the size of the $\sup W_T(\upsilon)$ under the null of a linear model. As mentioned above, there is an inherent arbitrariness in the choice of the weighting function, hence, we also report the results for $\text{ave} W_T(\upsilon)$ and $\text{exp} W_T(\upsilon)$ for the sake of completeness. Analyzing the size properties of these statistics involves generating data from the null model in (12) and the calculation of $p$–values for each simulated sample. Table 2 presents the rejection frequencies of the tests for different sample sizes.

We cannot attest that the $\sup W_T(\upsilon)$ test over-rejects compared with the other weighting functions used. In fact, the rejection frequencies are quite similar across the different weighting functions and clearly improve with increasing $T$.

In order to explore the power of the tests against the two-regime alternative, we generate data from a slightly modified process compared to (13). To take into account the strength of the
\[
\Delta y_t = \alpha \beta' y_{t-1} + \Gamma^A \Delta y_{t-1} + \delta (\Gamma^B - \Gamma^A) \Delta y_{t-1} I(z_{t-d} > \upsilon) + u_t.
\] (14)

If \(\delta = 0\), (14) corresponds to the null model in (12), and if \(\delta \neq 0\) the alternative of a two-regime model holds. The results of the power study for different sample sizes and several values for \(\delta\) are reported in Table 3.

<table>
<thead>
<tr>
<th>(supW_T(\upsilon))</th>
<th>T</th>
<th>(\delta = 0.2)</th>
<th>(\delta = 0.4)</th>
<th>(\delta = 0.6)</th>
<th>(\delta = 0.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.069</td>
<td>0.127</td>
<td>0.210</td>
<td>0.325</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.161</td>
<td>0.381</td>
<td>0.571</td>
<td>0.727</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.530</td>
<td>0.874</td>
<td>0.964</td>
<td>0.987</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(aveW_T(\upsilon))</th>
<th>T</th>
<th>(\delta = 0.2)</th>
<th>(\delta = 0.4)</th>
<th>(\delta = 0.6)</th>
<th>(\delta = 0.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.058</td>
<td>0.130</td>
<td>0.205</td>
<td>0.295</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.144</td>
<td>0.349</td>
<td>0.507</td>
<td>0.659</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.468</td>
<td>0.830</td>
<td>0.946</td>
<td>0.976</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(expW_T(\upsilon))</th>
<th>T</th>
<th>(\delta = 0.2)</th>
<th>(\delta = 0.4)</th>
<th>(\delta = 0.6)</th>
<th>(\delta = 0.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.077</td>
<td>0.135</td>
<td>0.214</td>
<td>0.331</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.167</td>
<td>0.393</td>
<td>0.577</td>
<td>0.724</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.530</td>
<td>0.874</td>
<td>0.974</td>
<td>0.990</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Size of the \(supW_T(\upsilon)\), \(aveW_T(\upsilon)\), and \(expW_T(\upsilon)\) tests

Table 3: Power of the \(supW_T(\upsilon)\), \(aveW_T(\upsilon)\), and \(expW_T(\upsilon)\) tests, 5% nominal size against two-regime alternative

As expected, the power of all statistics considered increases in the sample size and the threshold effect. We can confirm that the \(expW_T(\upsilon)\) statistic indeed has the highest rejection rate, as in Andrews in Ploberger (1994). However, compared with the other statistics considered, the difference is surprisingly small.
6 Empirical Illustration: Reinvestigating the Stability of the U.S. Money Demand

6.1 Motivation

Money demand constitutes a cornerstone to set up monetary policy. Hence, numerous empirical studies have investigated its determinants as well as its structural form (see Sriram, 2001 for an extensive survey). This paper proposes to shed new light on both of these issues.

Following (Lütkepohl 1993 or Ericsson, 1999), the standard form of the money demand is 
\[ M = f(Y, P, r, z), \]
where \( M \) is a money variable, \( Y \) a transaction volume, \( P \) a price level index, \( r \) an opportunity cost of holding money, and \( z \) a set of extra variables. Usually, there is a wide consensus regarding the series that are used for the three first determinants\(^9\), whereas divergences are observed for the other series. \( r \) should measure the opportunity cost of holding liquidity rather than investing it. Numerous interest rates on assets with different maturities have been employed\(^10\) leading to different results. Similarly there is no consensus on the composition of the set of extra variables. As mentioned by Choi and Jung (2009) the exchange rate regime, and in particular the introduction of the flexible exchange rates after the collapse of the Bretton woods system, may have modified the cash management practice. Thus, an effective nominal or real exchange rate can be introduced as in Rao and Kumar (2009). Friedman (1988) stresses the role played by the asset prices on real money holdings via the wealth and the substitution effect, suggesting to include an asset price\(^11\) as determinant of money holdings. Hence, it results that any conclusion concerning money demand should be carefully checked under different specifications of the demand function.

Particular attention has to be paid also to the structural form of the money demand. Since

\(^9\) The theory of money suggests to consider the most liquid form of money demand, i.e. the M1 aggregate, the transaction volume is either measured by GDP, GNP or GNI (see Bomberger and Makinen, 1980), while prices are either the implicit deflator of gdp or the consumer price index.

\(^10\) See Judd and Schadding (1982) for a survey. To sum up Goldfeld (1973) considers the money market rate and the interest rate on saving deposit, Hamburger (1977) the government bonds and the dividend price ratio on equity and Lütkepohl (1993) the discount rate on 91-day treasury bills.

\(^11\) Avouy-dovi et al. (2004) considers for the European money demand the real index is the ratio of the back projected Eurostoxx index over the GDP deflator.
are generally used to model money demand. They offer the advantage of clearly separating the long-run from the short-run behavior; where the first part covers economic theory and the latter relates to transitory effects. Several studies using a similar framework find money demand being quite stable in the post-war period\textsuperscript{12}, others conclude that it appears to be unstable after the 1970’s.\textsuperscript{13} The sources of this instability is often associated to the advent of financial innovations, to a permanent change in the relationship between money and interest rates, or other momentous (political) events. These instabilities can be incorporated in the linear ECM framework by introducing structural breaks at known or unknown dates.\textsuperscript{14} However, there is empirical evidence (see e.g. Lütkepohl, 1993 and 1999 or Escribano, 2004) that money demand instability should not be associated with long-run changes but should be instead related to modifications in the short-run dynamics. Furthermore, incorporating these transitory non-linearities would also lead to a better understanding of the episodes of missing money \textsuperscript{15}, of great velocity decline, which takes place in the early 80’s or of a liquidity trap (see Weberpals, 1997). In related studies, Sarno et. al (2003) and Calza and Zaghini (2009) investigate this class of non-linearity using a STR and a Markov-switching framework, respectively. Both studies indeed find a rather stable long-run behavior, but significant non-linearities in the short-run. In the following, we apply our SR-TVECM approach to the same research question, paying special attention on the relationship between the business cycle movements and money demand. That is, we analyze whether business cycle (reflecting the economy’s short-run behavior) movements cause instabilities in the demand for money.

6.2 Data and the Specification of the SR-TVECM

The empirical exercise consists of applying the methods elaborated in the foregoing sections. More precisely, we investigate whether U.S. money demand is non-linear across the business cycle, i.e. if (short-run) economic fluctuations govern monetary demand regimes or not. The threshold variable


\textsuperscript{13}See Goldfeld (1976), Judd and Scadding (1982), Friedman and Kuttner (1992) and Duca and van Hose (2004)

\textsuperscript{14}See e.g. the recent paper of Rao and Kummar (2009) or Lütkepohl et. al (1998).

\textsuperscript{15}It is simultaneous to the nomination of Paul Volcker at the head of the Federal Reserve.
which proxies the stance of the business cycle, is chosen as the coincident indicator proposed by the Chicago FED (CFNAI index), which corresponds to the index of economic activity developed by Stock and Watson (1999).

As mentioned earlier, there is a broad consensus that money demand is a function of the following arguments: the money variable, the transaction volume, the price level, the opportunity cost of holding money and some additional variables influencing money demand. Since none of the above mentioned variables exists per se we proxy them as follows.

- The Money variable is unambiguously the most liquid money aggregate, i.e. $M_1$.
- Transaction costs are proxied by GNP. It would have been possible to consider GDP as in Lütkepohl (1993), but GNP is more frequently used in related studies. Moreover, as Laidler (1993) shows, both variables do not present major differences.
- Considering the price level, the GNP deflator is consequently the straightforward counterpart.
- As noticed earlier, proxying the opportunity cost of holding money is more difficult and several interest rates are used in the literature. We will consider four of them: The Federal Funds rate, the short-term rate on commercial paper, the short-term deposit rate, and the 3-month treasury bill. They cover the different maturities used in related studies.
- To complete the money demand system, we consider two additional variables: The real effective exchange rate and the Dow Jones Industrial Average Index to control for competitiveness and financial effects.

The period considered ranges from 1973 until 2009. All series are in monthly frequency except the GNP and the GNP deflator which are quarterly. We use a method based on Kalman filtering.

\[\text{\footnotesize 16 see for a description: http://www.chicagofed.org/webpages/publications/cfnai/index.cfm}\]
to interpolate these series to monthly data. All data is freely available on the internet.

In order to draw robust conclusions with respect to the constitution of the money demand function, we will consider any possible combination of the above mentioned variables. That is, we estimate and test for short-run non-linearities in a VECM using 16 different specifications of data sets. All variables are taken in logs of levels except the interest rates and the CFNAI. Every data set contains the GNP, the GNP deflator, and the money stock M1. Additional variables are specified in Table 4.

| Data Set 1 | Federal Funds rate |
| Data Set 2 | Deposit rate |
| Data Set 3 | Commercial paper rate |
| Data Set 4 | T-Bill |
| Data Set 5 | Federal Funds rate | Exchange rate |
| Data Set 6 | Deposit rate | Exchange rate |
| Data Set 7 | Commercial paper rate | Exchange rate |
| Data Set 8 | T-Bill | Exchange rate |
| Data Set 9 | Federal Funds rate | Dow Jones Index |
| Data Set 10 | Deposit rate | Dow Jones Index |
| Data Set 11 | Commercial paper rate | Dow Jones Index |
| Data Set 12 | T-Bill | Dow Jones Index |
| Data Set 13 | Federal Funds rate | Exchange rate | Dow Jones Index |
| Data Set 14 | Deposit rate | Exchange rate | Dow Jones Index |
| Data Set 15 | Commercial paper rate | Exchange rate | Dow Jones Index |
| Data Set 16 | T-Bill | Exchange rate | Dow Jones Index |

Table 4: Definition of different data sets.

\footnote{In fact, we use the interpolation procedure proposed by Bernanke, Gertler and Watson (1997) in a version as implemented in Mönch and Uhlig (2005). The real GNP is interpolated to monthly data by using industrial production, the civilian unemployment rate and real disposable income, whereas the GNP deflator is interpolated using the Consumer and the producer Price Index.}

\footnote{See appendix for data sources.}
For the specification of each (data set specific) SR-TVECM (resp. VECM) we use an extended version of the BIC to determine the lag-order \( p \) and the delay-parameter \( d \) of the threshold variable, while the threshold process itself is well described by an AR(2) process with iid \( WN \) residuals. The trimming value is set to 0.1. While testing for threshold-effects we use 2,000 bootstrap replications to approximate the distribution of the test statistics.

### 6.3 Estimation and Testing Results

The estimation results are reported in Table 5, i.e. the model specification, the estimated threshold value, and the number of observations in each regime for each data set considered. As it turns out that linearity is strongly rejected (at the 1% level) for every specification of the underlying money demand system, p-values are not reported. This high rejection frequency indicates surprisingly strong threshold effects in the short-run money demand, confirming hence the findings of Lütkepohl (1993) and others.

<table>
<thead>
<tr>
<th>Data Set 1</th>
<th>Lag-order</th>
<th>Delay parameter</th>
<th>Threshold value</th>
<th>Obs. in lower regime</th>
<th>Obs. in upper regime</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>-1.09</td>
<td>51</td>
<td>384</td>
</tr>
<tr>
<td>Data Set 2</td>
<td>4</td>
<td>4</td>
<td>-1.09</td>
<td>51</td>
<td>384</td>
</tr>
<tr>
<td>Data Set 3</td>
<td>4</td>
<td>4</td>
<td>-1.09</td>
<td>51</td>
<td>384</td>
</tr>
<tr>
<td>Data Set 4</td>
<td>4</td>
<td>4</td>
<td>-1.09</td>
<td>51</td>
<td>384</td>
</tr>
<tr>
<td>Data Set 5</td>
<td>4</td>
<td>2</td>
<td>-0.83</td>
<td>72</td>
<td>368</td>
</tr>
<tr>
<td>Data Set 6</td>
<td>3</td>
<td>2</td>
<td>-0.83</td>
<td>72</td>
<td>369</td>
</tr>
<tr>
<td>Data Set 7</td>
<td>3</td>
<td>2</td>
<td>-0.83</td>
<td>72</td>
<td>369</td>
</tr>
<tr>
<td>Data Set 8</td>
<td>4</td>
<td>2</td>
<td>-0.83</td>
<td>72</td>
<td>368</td>
</tr>
<tr>
<td>Data Set 9</td>
<td>3</td>
<td>2</td>
<td>-1.24</td>
<td>44</td>
<td>397</td>
</tr>
<tr>
<td>Data Set 10</td>
<td>3</td>
<td>2</td>
<td>-1.24</td>
<td>44</td>
<td>397</td>
</tr>
<tr>
<td>Data Set 11</td>
<td>3</td>
<td>2</td>
<td>-1.24</td>
<td>44</td>
<td>397</td>
</tr>
<tr>
<td>Data Set 12</td>
<td>3</td>
<td>2</td>
<td>-1.24</td>
<td>44</td>
<td>397</td>
</tr>
<tr>
<td>Data Set 13</td>
<td>4</td>
<td>2</td>
<td>-0.83</td>
<td>72</td>
<td>368</td>
</tr>
<tr>
<td>Data Set 14</td>
<td>3</td>
<td>2</td>
<td>-0.92</td>
<td>66</td>
<td>375</td>
</tr>
<tr>
<td>Data Set 15</td>
<td>3</td>
<td>2</td>
<td>-0.92</td>
<td>66</td>
<td>375</td>
</tr>
<tr>
<td>Data Set 16</td>
<td>4</td>
<td>2</td>
<td>-0.83</td>
<td>72</td>
<td>368</td>
</tr>
</tbody>
</table>

Table 5: Estimation results for all data sets considered.

The timing of money demand regimes are reported in Figure 1. It is noticeable that the dating of the regimes is very similar for all systems considered. A simple correlation analysis between the system indicates that the regimes are synchronized by more than 80%. Such a consensus result constitutes a novelty for this literature, where the conclusion are often contradicting each others, depending on the composition of the system. Besides, our results also support they ideas that
money demand is not only suffering from structural breaks, but are driven by regimes governed by the stance of the economy. Looking more closely at the regime we obtain, it appears that our SR-TVECM identifies the episodes of missing money already extensively described in the literature (mid 70’s and early 80’s). It is also noticeable that such a regime also occurs in the early 2000’s and since the beginning of the sub-prime crisis. Nevertheless, these episodes (i.e. those associated with an economic activity contraction) are not numerous (between 10 and 15) and are on average five time shorter than the “normal” states. Interestingly, this low money demand regime is prevailing since 2007, which is quite long (2 years) compared to its historical predecessors, stressing hence the deepness of the current financial crisis.

7 Conclusion

This paper develops an estimation and testing framework for a threshold VECM, where short-run dynamics are regime dependent and are driven by an exogenous, stationary and ergodic threshold variable. We modify the traditional Wald-type test for linearity and derive its asymptotic distribution, which turns out to be non-standard, but similar to the one presented in Andrews (1993).

We apply this approach to investigate the stability of money demand in the U.S. In order to account for the diversity of the potential determinants of money holdings, a broad scope of money demand systems is analyzed. The implementation of the linearity test allows us to confirm that money demand is not only suffering from structural breaks, but is driven by regimes governed by the stance of the economy. Such results have strong consequences for modelers and consequently for central bankers in conducting monetary policy.
8 Appendix

8.1 Definitions of the variables

If not indicated differently the following series are taken from the Federal Reserve Board of St. Louis’ web site http://research.stlouisfed.org/fred2/.

**GNP**: Is the real gross national product in chained 2005 dollars which is the series \( GNPC96 \).

**M1**: Is the M1 money stock which is the series \( M1SL \).

**GNP Deflator**: The implicit price deflator of the GNP is the series \( GNPDEF \).

**Federal Funds rate**: The effective federal funds rate \( FEDFUNDS \).

**Short-term rate on commercial paper**: 1-Month Nonfinancial Commercial Paper Rate \( CP1M \) (1973-1997), \( CPN1M \) (1998-2009).

**Short-term deposit rate**: The 1-Month Certificate of Deposit: Secondary Market Rate \( CD1M \).

**3-month Treasury Bill**: 3-Month Treasury Bill: Secondary Market Rate \( TB3MS \).

**Industrial production**: \( INDPRO \)

**Civilian unemployment rate**: \( UNRATE \)

**Real disposable personal income**: \( DSPIC96 \)

**Consumer price index**: \( CPIAUCSL \)

**Producer price index**: Finished goods, \( PPIFGS \)

**Real effective exchange rate**: The real price-adjusted broad dollar index, available online from the Board of Governors of the Federal Reserve’s web site (http://www.federalreserve.gov/).

**Dow Jones Industrial Average Index**: In monthly adjusted close prices available online from finance.yahoo.com.
Some of the proofs presented below rely on the following uniform strong law.

**Lemma 8** Let the threshold variable \( \{z_t\} \) be strictly stationary and ergodic with continuous distribution function \( F \). Further, assume that for the data \( \{y_t\} \) (stationary or not) \( E|f(y_t)| < \infty \) holds, where \( f \in \mathcal{F} \) with \( \mathcal{F} \) being a class of (measurable) functions on a set with a sigma-field that carries a probability measure \( P \), then

\[
\sup_{v \in \mathcal{V}} \left| P_T f(y_t|z_t \leq v) - P f(y_t) F(v) \right| \to 0 \quad \text{almost surely} \quad (15)
\]

as \( T \to \infty \), where the empirical measure \( P_T \) is constructed by sampling from \( P \).

**Proof.** The proof is based on the direct approximation theorem as presented in Pollard (1984, Thm. II/2). We have to show that \( \mathcal{F} \) can be approximated by a finite class of functions \( \{f^u_{\epsilon,k}(y_t), f^l_{\epsilon,k}(y_t) : k = 1, \ldots, K\} \) which have the property that for each \( v \) there exist some \( k \) such that \( f^u_{\epsilon,k}(y_t) \leq f(y_t|z_t \leq v) \leq f^l_{\epsilon,k}(y_t) \) and \( E|f^u_{\epsilon,k}(y_t) - f^l_{\epsilon,k}(y_t)| < \epsilon \) hold for any \( \epsilon > 0 \). If \( \{z_t\} \) is strictly stationary and ergodic with continuous distribution function \( F \) and \( E|f(y_t)| < \infty \), we can construct a suitable set of approximation functions in a similar way as in Hansen (1996, Lemma 1).

Before proofing Lemma 2, we establish another result we will need below.

**Lemma 9** If \( u_t \) is any real-valued martingale difference sequence and \( \{z_t\} \) is stationary and ergodic with continuous distribution function \( F \), and \( y_t \) follows an \( I(1) \) process as defined in (2), and the initial values are assumed to be zero w.l.o.g., then the following results hold.

1. The process \( y_T \) can be written as

\[
y_T = \Xi^A \sum_{t=1}^{T} u_t I(z_{t-d} \leq v) + \Xi^B \sum_{t=1}^{T} u_t I(z_{t-d} > v) + \text{stationary part} \quad (16)
\]

where \( \Xi^m = \beta_{\perp} \left[ \alpha_{\perp} \left( J_K - \sum_{i=1}^{p-1} \Gamma^m_i \right) \beta_{\perp} \right]^{-1} \alpha'_{\perp} \), with \( \alpha_{\perp} \) and \( \beta_{\perp} \) being the usual orthogonal
complements of the reduced rank matrices $\alpha$ and $\beta$, and for $m = A, B$.

2. The asymptotic properties of the process $\{y_T\}$ can be summarized as

$$
\frac{y_{[T]}}{\sqrt{T}} = \Xi^A \frac{1}{\sqrt{T}} \sum_{t=1}^{[T]} u_t I(z_{t-d} \leq v) + \Xi^B \frac{1}{\sqrt{T}} \sum_{t=1}^{[T]} u_t I(z_{t-d} > v) + \underbrace{\text{stationary part}}_{\to 0} \sqrt{T} (17)
$$

and where

- $\Xi^A \frac{1}{\sqrt{T}} \sum_{t=1}^{[T]} u_t I(z_{t-d} \leq v) \overset{d}{\to} F(v) \Xi^A W(r)$
- $\Xi^B \frac{1}{\sqrt{T}} \sum_{t=1}^{[T]} u_t I(z_{t-d} \leq v) \overset{d}{\to} (1 - F(v)) \Xi^B W(r)$

as $T \to \infty$. $W(r)$ denotes a standard Wiener process.

Proof. Since the process is piecewise linear and the transmission from one regime to another is exogenously determined by the threshold process with the above mentioned properties, we can adapt the proof in Lütkepohl (2006, Prop. 6.1) to each regime separately and use Lemma 8 to verify the above statements. ■

8.2 Proof of Lemma 2

To show asymptotic equivalence we have to prove that $u_t - u_t^* \overset{p}{\to} 0$. For this purpose, note first that $u_t^*$ are just the residuals resulting from regressing $u_t$ on $y_{t-1}$. Hence,

$$
u_T - u_T^* = \sum_{t=1}^{T} u_t y_{t-1}^j \left( \sum_{t=1}^{T} y_{t-1} y_{t-1}^j \right)^{-1} y_{T-1} = 1 \sum_{t=1}^{T} u_t y_{t-1}^j \left( \frac{1}{T} \sum_{t=1}^{T} y_{t-1} y_{t-1}^j \right)^{-1} \frac{1}{T} y_{T-1}. (18)$$

In order to study the limit behavior of this equation notice that if the system is $I(1)$, $\sum u_t y_{t-1} = O_p(T)$ and $\sum y_{t-1} y_{t-1}^j = O_p(T^2)$ holds. From Lemma 9 we have that $\frac{y_{T-1}}{\sqrt{T}} = O_p(1)$ and hence it follows that $\frac{y_{T-1}}{T} = o_p(1)$. Thus, $u_T - u_T^* = O_p(1)O_p(1) = o_p(1)$. ■

However, even if $\text{plim} \left( \frac{Y_{t-1} Y_{t-1}}{T^2} \right)$ exists, it might be a singular matrix as $T \to \infty$ because of the co-integration among the variables. To overcome this problem, the above cross-product can be decomposed into non-singular components (see Ahn and Reinsel, 1990), such that the existence of the inverse is guaranteed.
8.3 Proof of Lemma 5:

In order to simplify notation we can write the multiple time series \( y_t \) in matrix form as \( Y = Y Y'_{-1} (Y_{-1} Y'_{-1})^{-1} Y_{-1} \). Now, it is directly seen that if \( \text{plim} \left( \frac{Y Y'_{-1}}{T} \right) = \text{plim} \left( \frac{Y_{-1} Y'_{-1}}{T} \right) \), \( Y \) would converge to \( \Delta Y \) as \( T \to \infty \). Assuming \( y_t \) is \( I(1) \) we know that \( \left( \frac{Y_{-1} Y'_{-1}}{T} \right) = O_p(1) \). Using simple algebra we find that

\[
Y Y'_{-1} - Y_{-1} Y'_{-1} = \sum y_t y'_t - \sum y_{t-1} y'_{t-1} = \sum \Delta y_t y'_{t-1} = \Delta Y Y'_{-1} = O_p(T).
\]

(19)

Now scaling by \( T^2 \) and taking the probability limit we have that \( \text{plim} \left( \frac{Y_{-1} Y'_{-1}}{T^2} \right) = \text{plim} \left( \frac{Y Y'_{-1}}{T^2} \right) - o_p(1) \).

8.4 Derivation of the test statistic

For this purpose it is convenient to write the model (4) in matrix notation:

\[
\Delta \hat{Y} = \Gamma^* \hat{X}^* + U,
\]

(20)

where \( \Gamma^* = [\Gamma^A, \Gamma^B] \) and \( \hat{X}^* = [X^A(\upsilon), X^B(\upsilon)]' \). For later use we introduce \( \gamma^* = \text{vec}(\Gamma^*) \), \( \gamma^A = \text{vec}(\Gamma^A) \), \( \gamma^B = \text{vec}(\Gamma^B) \), \( R = [I_{K^2(p-1)}, -I_{K^2(p-1)}] \). Since we have established asymptotic normality of the residuals as well as asymptotic stationarity of the regressors, we can conclude that \( \text{plim} \left( \frac{X^A(\upsilon)(X^A(\upsilon))'}{T} \right) \) and \( \text{plim} \left( \frac{X^B(\upsilon)(X^B(\upsilon))'}{T} \right) \) exist and are non-singular. Obviously we have \( \text{plim} \left( \frac{X^A(\upsilon)(X^B(\upsilon))'}{T^2} \right) = 0 \) due to the indicator function. Using simple algebra a standard Wald
test statistic can now be written as:

\[ W_T(v) = T \gamma^* R' \left( R \left( \frac{XX'}{T} \right)^{-1} \otimes \hat{\Sigma}_u \right) R' \right)^{-1} R^* \]

\[ = T(\gamma^A - \gamma^B)^y \left( R \left( \frac{XX'}{T} \right)^{-1} \otimes \hat{\Sigma}_u \right) R' \right)^{-1} (\gamma^A - \gamma^B) \]

\[ = T(\gamma^A - \gamma^B)^y \left( R \left( \begin{bmatrix} \hat{X}J^A(u)X^A(u)' \hat{X}J^A(u)X^A(u)' \\ \hat{X}J^B(u)X^B(u)' \hat{X}J^B(u)X^B(u)' \end{bmatrix}^{-1} \otimes \hat{\Sigma}_u \right) R' \right)^{-1} (\gamma^A - \gamma^B) \]

\[ = T(\gamma^A - \gamma^B)^y \left( R \left( \begin{bmatrix} \hat{X}J^A(u)X^A(u)' & 0 \\ 0 & \hat{X}J^B(u)X^B(u)' \end{bmatrix}^{-1} \otimes \hat{\Sigma}_u \right) R' \right)^{-1} (\gamma^A - \gamma^B) \]

\[ = T(\gamma^A - \gamma^B)^y \times \left\{ \frac{\hat{X}J^A(u)X^A(u)'}{T} \left( \frac{XX'}{T} \right)^{-1} \frac{\hat{X}J^A(u)X^A(u)'}{T} \otimes \hat{\Sigma}_u^{-1} \right\} \times (\gamma^A - \gamma^B), \]

with \[ \left( \frac{XX'}{T} \right)^{-1} = \left( \frac{\hat{X}J^A(u)X^A(u)'}{T} + \hat{X}J^B(u)X^B(u)' \right)^{-1}. \] Now, with \[ \hat{X}^A(u) := \hat{X}J^A(u) \] and \[ \hat{X}^B(u) := \hat{X}J^B(u), \] we get the statistic presented in the text.

### 8.5 Proof of Proposition 6:

**Lemma 10** Under \( \mathcal{H}_0: \sup_{v \in \Upsilon} W_T(v) \Rightarrow \sup_{v \in \Upsilon} \psi(v)'(\Theta(v)\psi(v), \text{where} \ \Theta(v) := \left[ Q \otimes \Sigma_u F(v)(1 - F(v)) \right]^{-1} \) and \( \psi(v) \) is a zero-mean gaussian matrix with conditional covariance \( \min\{F(v_i), F(v_j)\}(1 - \min\{F(v_i), F(v_j)\}) \times Q \otimes \Sigma_u). \)

**Proof.** The proof follows similar arguments as Prop.1 in Gonzalo and Pitarakis (2007). It follows from Lemma 2 and Lemma 5 that \( Q := \text{plim} \left( \frac{XX'}{T} \right) \) exists and is non-singular. From Lemma 8 it follows that \( \frac{\hat{X}^A(u)X^A(u)'}{T} \rightarrow_{p} F(v)Q \) and similarly \( \frac{\hat{X}^B(u)X^B(u)'}{T} \rightarrow_{p} (1 - F(v))Q, \) given Ass.1 holds. Since the regressors in (5) are stationary, \( vec(U\hat{X}^{A,B}') \) is a matrix of stacked vector martingale difference sequences under the stated conditions about the innovations. Thus, we can easily establish the consistency of the parameter estimates: \( \text{plim}(\Gamma^A - \Gamma^A) = \text{plim} \left( \frac{UX^A(u)'}{T} \right) \times \text{plim} \left( \frac{XX'}{T} \right)^{-1} = 0 \times F(v)Q = 0 \) and \( \text{plim}(\Gamma^B - \Gamma^B) = \text{plim} \left( \frac{UX^B(u)'}{T} \right) \times \text{plim} \left( \frac{XX'}{T} \right)^{-1} = 0 \times (1 - F(v))Q = 0. \)
Since \( \hat{\Gamma}^m \mathop{\rightarrow}^d \Gamma^m \) implies \( \hat{\Gamma}^m \mathop{\rightarrow}^m \Gamma^m \), \( m = A, B \), we can use standard arguments (e.g. Lütkepohl (2006), prop. 3.2) to show that \( \hat{\Sigma} \mathop{\rightarrow}^d \Sigma \). Collecting these results, we can show that

\[
\frac{X^B(v)X^B(v)'}{T} = \left( \frac{X^A(v)X^A(v)'}{T} \right)^{-1} \frac{X^A(v)X^A(v)'}{T} \mathop{\rightarrow}^d \hat{\Sigma}_u^{-1} F(v)(1 - F(v))Q \otimes \Sigma_u^{-1}. \tag{21}
\]

It remains to study the probability limit of \( \sqrt{T}(\hat{\gamma}^A - \hat{\gamma}^B) \). Note that standard LS-algebra yields

\[
\sqrt{T}(\hat{\gamma}^m - \gamma^m) = \left( \frac{X^m(v)X^m(v)'}{T} \right)^{-1} I_K = \frac{1}{\sqrt{T}} (X^m(v) \otimes I_K) vec(U), m = A, B. \]

Now, given that \( \gamma^A = \gamma^B \) under the null, we can use previous results to show

\[
\sqrt{T}(\hat{\gamma}^A - \hat{\gamma}^B) = (F(v)Q \otimes I_K)^{-1} \frac{1}{\sqrt{T}} \left( \frac{X^A(v)X^A(v)'}{T} \otimes I_K \right) vec(U) - (1 - F(v))Q \otimes I_K)^{-1} \frac{1}{\sqrt{T}} \left( \frac{X^B(v)X^B(v)'}{T} \otimes I_K \right) vec(U)
\]

\[
= (F(v)(1 - F(v))Q \otimes I_K)^{-1} \times \Psi(v), \tag{22}
\]

where \( \Psi(v) := (1 - F(v)) \frac{1}{\sqrt{T}} \left( \frac{X^A(v)X^A(v)'}{T} \otimes I_K \right) vec(U) - F(v) \frac{1}{\sqrt{T}} \left( \frac{X^B(v)X^B(v)'}{T} \otimes I_K \right) vec(U) \). To establish the distributional convergence properties of \( \Psi(v) \), note that for each \( v \in T, \hat{\gamma}_t^A(u_t) \) and \( \hat{\gamma}_t^B(u_t) \) are sequence stable stationary martingale differences, to which the pointwise central limit theorem applies. Hence, \( \frac{1}{\sqrt{T}} (X^B(v) \otimes I_K) vec(U) \mathop{\rightarrow}^d \mathcal{N}(0, (1 - \min\{F(v_i), F(v_j)\})Q \otimes \Sigma_u) \) and \( \frac{1}{\sqrt{T}} (X^B(v) \otimes I_K) vec(U) \mathop{\rightarrow}^d \mathcal{N}(0, (1 - \min\{F(v_i), F(v_j)\})Q \otimes \Sigma_u) \) for \( i \neq j \). Thus,

\[
\Psi(v) = (1 - F(v)).\mathcal{N}(0, \min\{F(v_i), F(v_j)\})Q \otimes \Sigma_u) - F(v).\mathcal{N}(0, (1 - \min\{F(v_i), F(v_j)\})Q \otimes \Sigma_u)
\]

\[
= \mathcal{N}(0, (\min\{F(v_i), F(v_j)\}(1 - \min\{F(v_i), F(v_j)\}) \times Q \otimes \Sigma_u). \tag{23}
\]

Now, by combining (21), (22), and (23) we can establish the asymptotic distribution of \( W_T(v) \), namely

\[
W_T(v) \Rightarrow \left[ (F(v)(1 - F(v))Q \otimes I_K)^{-1} \times \Psi(v) \right] \times F(v)(1 - F(v))Q \otimes \Sigma_u^{-1}
\]

\[
\times \left[ (F(v)(1 - F(v))Q \otimes I_K)^{-1} \times \Psi(v) \right]
\]

\[
\Rightarrow \Psi(v)'\Theta(v)\Psi(v), \tag{24}
\]

25
where $\Theta(v) := \left[ Q \otimes \Sigma_u F(v)(1 - F(v)) \right]^{-1}$ as $T \to \infty$. The result stated in Lemma 10 follows now directly from the continuous mapping theorem under suitable conditions (see: Remark 13).

**Remark 11** The limiting distribution of $\text{AveWald}_T(v)$ and $\text{ExpWald}_T(v)$ also follows from (24) by applying the continuous mapping theorem.

**Corollary 12** The asymptotic distribution of $W_T(v)$ derived in Lemma 6 is similar to a squared $K^2(p - 1)$-dimensional standardized tied-down Bessel process that appears in Andrews (1993).

**Proof.** To simplify notation let $s := F(v) \in S$ with $S := [F(v_L), F(v_U)] \subset (0, 1)$. The result follows directly from Lemma 10 by re-writing (24) as

$$W_T(s) \Rightarrow \frac{\Psi^*(s)\Psi^*(s)}{s(1 - s)} =: \Omega(s),$$

where $\Psi^*(s) := \mathcal{N}(0, \min\{s_i, s_j\}(1 - \min\{s_i, s_j\}))I_{K^2}$ is similar to a $K^2(p - 1)$-dimensional Brownian bridge on $[0, 1]$ restricted to $S$.

**Remark 13** As shown in Andrews (1993) Cor. 1, the continuous mapping theorem only leads to the desired result, if $S$ has closure in $(0, 1)$, which is fulfilled by Ass. 1.
References


Figure 1: Timing of identified money demand regimes.