UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Center For International Economics

Working Paper Series

Working Paper No. 2011-11

Data-driven estimation of diurnal duration patterns

Yuanhua Feng

December 2011

Center For International Economics

Data-driven estimation of diurnal duration patterns

Yuanhua Feng

University of Paderborn

Abstract

This paper proposes a local linear estimator for diurnal patterns of transaction durations under a special nonparametric regression model, whose asymptotics are different to any known results. An iterative plug-in algorithm is developed for selecting the bandwidth. The ACD model is then applied to analyze the standardized durations. Data examples show that the proposals work well in practice.

Keywords: Autoregressive conditional duration, diurnal duration patterns, local linear estimator, bandwidth selection, iterative plug-in.

JEL Codes: C14, C41

1 Introduction

Let $t_0 < t_1 < ... < t_N < t_{N+1} = T$ be the time points at which trades occur and $x_i = t_i - t_{i-1}$ the duration between trades, where $t_0 = 0$ and N is the (random) number of trades. Consider the model $x_i = \phi_i \psi_i \varepsilon_i$, where ϕ_i is a (deterministic) diurnal pattern and ψ_i is the conditional mean after removing ϕ_i. Estimation of $\phi(t)$ is a necessary step for further econometric analysis of transaction durations using e.g. the ACD (autoregressive conditional duration) of Engle and Russell (1998). Different approaches for estimating $\phi(t)$ are introduced in the literature. For instance, Engle and Russell (1998) propose to use a cubic spline and a nonparametric estimator was introduced by Veredas et al. (2002). For further ideas on the estimation of ϕ see e.g. Bauwens and Giot (2000), Hautsch (2004) and references therein.

This paper focuses on developing a data-driven local linear estimator of ϕ. For this purpose a nonparametric regression model with some special features is defined, where ϕ is the mean function and the scale function at the same time, and is also proportional to the reciprocal of the design density. The asymptotic variance and asymptotically optimal bandwidth of $\hat{\phi}$ are hence different to any known results. An iterative plug-in (IPI, Gasser et al., 1991) algorithm with some improvements is developed for selecting the bandwidth under the current model. Data examples show that the proposals work well in practice.

2 Local linear estimation of the diurnal pattern

The model $x_i = \phi_i \psi_i \varepsilon_i$ (Engle and Russell, 1998, 2010) can be rewritten as follows:

$$x_i = \phi(t_i) + \phi(t_i)(y_i - 1), i = 1, ..., N, \tag{1}$$

where $\phi(t_i)$ is the local mean of x_i, $y_i = \psi_i \varepsilon_i$ with $E(y_i) = 1$, ψ_i is the conditional mean of y_i, which can be modeled by a unit ACD with $\psi_i = \omega + \sum_{j=1}^{p} \alpha_j y_{i-j} + \sum_{k=1}^{q} \beta_k \psi_{i-k}$, $\omega > 0$, $\alpha_j, \beta_k \geq 0$ and $\omega = 1 - \sum \alpha_j - \sum \beta_k$, and ε_i are i.i.d. non-negative random variables with $E(\varepsilon_i) = 1$. Let $\xi_i = y_i - 1$. Model (1) is a special nonparametric regression with stationary errors ξ_i. 1: $\phi(t)$ is its mean and scale function at the same time; 2. $x_i = t_i - t_{i-1}$ is determined by the regressor and 3. $\phi(t_i) \to 0$, as $N \to \infty$. Moreover, let $f(t) > 0$ be the design density of t and $m(t) = [f(t)]^{-1}$, we have $N\phi(t_i) = m(t_i) + o(N^{-1})$. These features will affect the asymptotic properties of $\hat{\phi}(t)$ strongly. We see $m(t)$ is a given function, but $\phi(t)$ depends on N. Hence we will first estimate $m(t)$ from $z_i = Nx_i$ and then put $\hat{\phi}(t) = \hat{m}(t)/N$, since the discussion with $\hat{m}(t)$ is more convenient.

The derivatives $m^{(\nu)}(t)$ can be estimated by minimizing the weighted least squares

$$Q = \sum_{i=1}^{N} \{z_i - a_0(t) - a_1(t)(t_i - t) - ... - a_p(t)(t_i - t)^p\}^2 K\left(\frac{t_i - t}{b}\right), \tag{2}$$

where K is a kernel and b is the bandwidth. We have $\hat{m}^{(\nu)}(t) = \nu! \hat{a}_\nu$ and $\hat{\phi}^{(\nu)}(t) = \nu! \hat{a}_\nu / N$. In the sequel, local linear estimators $\hat{m}(t)$ and $\hat{\phi}(t) = \hat{m}(t)/N$ will be used. Let $\hat{y}_i = x_i / \hat{\phi}(t_i)$. A suitable ACD model can be fitted to $\hat{y}_i$ using the standard QML method.

Let $\text{var}\,[\hat{m}(t)]$, $B[\hat{m}(t)]$, $\text{var}\,[\hat{\phi}(t)]$ and $B[\hat{\phi}(t)]$ denote the variance and bias of $\hat{m}(t)$ and $\hat{\phi}(t)$, respectively. We have $\text{var}\,[\hat{\phi}(t)] = \text{var}\,[\hat{m}(t)]/N^2$ and $B[\hat{\phi}(t)] \approx B[\hat{m}(t)]/N$. Asymptotic properties of $\hat{m}(t)$ can be derived under regularity conditions. The asymptotic bias of $\hat{m}(t)$ is the same as for a common local linear estimator. Let $\gamma(k) = \text{cov}\,(y_i, y_{i+k})$, $S = \sum \gamma(k)$ and $R(K) = \int K^2(u)du$. By adapting known results in the literature, we can obtain the asymptotic variance of $\hat{m}(t)$ at $0 < t < T$:

$$\operatorname{var}[\hat{m}(t)] \approx \frac{R(K)S}{Nbf(t)} m^2(t) = \frac{R(K)S}{Nb} m^3(t), \tag{3}$$

which is proportional to $m^3(t)$. Furthermore, let $I(K) = \int u^2 K(u) du$. The optimal bandwidth for estimating $m(t)$ and $\phi(t)$ on $[0, T]$, b_{opt}, is the same and is asymptotically

$$b_{\text{opt}} \approx b_{\text{A}} = \left(\frac{R(K)S}{I^2(K)} \frac{\int m^3(t) dt}{\int [m''(t)]^2 dt} \right)^{1/5} N^{-1/5}. \tag{4}$$

Let $V = R(K)S \int m^3(t) dt$ and $b = o(b_{\text{A}})$. At an interior point $b < t < T - b$, we have

$$\sqrt{Nb}[\hat{m}(t) - m(t)] \xrightarrow{\mathcal{D}} N(0, V). \tag{5}$$

3 Bandwidth selection

An IPI algorithm based on (4) with a starting bandwidth $b_0 = T/10$ is developed, which extends the idea of Gasser et al. (1991) to random design nonparametric regression with dependent errors, huge sample size and very large error variance, where S is estimated by a lag-window estimator with the maximal lag $K \approx 4N^{1/3}$ (see Bühlmann, 1996 and Feng and Wang, 2011). For any reasonable b_0, the IPI bandwidth converges to the same fix-point. The above choice of b_0 ensures that S can already be well estimated at the beginning. By an IPI algorithm, $m''(t)$ in the jth iteration is estimated using a bandwidth b_{2j} calculated from b_{j-1}. Here, the formula $b_{2j} = T(b_j/T)^{1/2}$ (Beran and Feng, 2002) is used, which works better than $\tilde{b}_{2j} = b_{j-1} * N^{1/10}$ (Gasser et al., 1991) in the current case. The choice of b_0 and b_{2j} also helps to reduce a few iterations. Starting with b_0, the algorithm reads as follows:

i) In the jth iteration, obtain $\hat{\phi}_j(t)$ with b_{j-1}. Let $\hat{y}_{ji} = x_i/\hat{\phi}_j(t_i)$. Calculate $\hat{\gamma}_j(k)$ from $\hat{y}_{ji}$ and let $\hat{S}_j = \sum_{|k|<K} w_k \hat{\gamma}_j(k)$, where w_k are the Bartlett-weights and $K \approx 4N^{1/3}$.

ii) Obtain $\hat{m}_j(t_i)$ with b_{j-1} and $\hat{m}_j''(t_i)$ using local cubic with $b_{2j} = T(b_{j-1}/T)^{\alpha}$, where $\alpha = 1/2$ or $\alpha = 5/7$. Let $\hat{I}_j(m^3) = \frac{1}{N} \sum_{i=1}^{N} [\hat{m}_j(t_i)]^3$ and $\hat{I}_j([m'']^2) = \frac{1}{N} \sum_{i=1}^{N} [\hat{m}_j''(t_i)]^2$.

iii) Improve b_{j-1} by $b_j = \left(\frac{R(K)\hat{S}_j}{I^2(K)} \frac{\hat{I}_j(m^3)}{\hat{I}_j([m'']^2)} \right)^{1/5} N^{-1/5}$ and increase j by one.

iv) Repeatedly carry out i) to iii) until convergence is reached and put $\hat{b}_A = b_j$.

Under regularity assumptions, in particular that $E(y_i^4) < \infty$, the rate of convergence of $\hat{b}_\mathrm{A}$ is the same as for nonparametric regression with independent errors and is determined by α. If $\alpha = 1/2$, $\hat{b}_\mathrm{A}$ is most stable so that $\hat{b}_\mathrm{A} \approx b_\mathrm{opt}[1+O(N^{-1/5})+O_p(N^{-1/2})]$. For $\alpha = 5/7$, $\hat{b}_\mathrm{A}$ achieves the highest rate of convergence with $\hat{b}_\mathrm{A} \approx b_\mathrm{opt}[1+O(N^{-2/7})+O_p(N^{-2/7})]$ (see Lemma 1 of Beran and Feng, 2002 with $\delta = 0$). Theoretically, the use of $\alpha = 2/7$ is more preferable, but the error variance in the current case is very large and the stability of $\hat{b}_\mathrm{A}$ is more important. Hence $\alpha = 1/2$ will be used in this paper. Furthermore, we propose to calculate the final estimates using the local bandwidths $\hat{b}(t) = \hat{b}_\mathrm{A} * \hat{m}(t)$. This is a quasi k-NN method and leads to further improvement of the estimation quality.

4 Practical implementation and applications

An R code is written by means of the "lm" function with proper weights, which is applied to stock durations of a few German firms ($T = 8.5$ and $b_0 = 0.85$ hours). It shows that the algorithm works well in practice. Bandwidths (in hours) selected with the bisquare kernel for BMW stock durations in the first two weeks in Aug 2011 are given in Table 1 as examples. Note that $\hat{b}_\mathrm{A}$ is jointly determined by N, S and the complexity of $\phi(t)$

Table 1. Selected bandwidths (h) for BMW diurnal duration patterns in two weeks

	1.8.11	**2.8.11**	3.8.11	4.8.11	5.8.11	**8.8.11**	9.8.11	10.8.11	11.8.11	12.8.11
N	9249	**14527**	12789	19477	22317	**24377**	34286	25995	19437	15404
$\hat{b}_A$	1.0657	**1.4086**	0.9794	0.8611	1.0867	**0.5565**	1.3248	0.8810	0.8397	0.9602

and changes strongly from one day to another. The largest bandwidth is $\hat{b}_\mathrm{A} = 1.4086$ on Aug 2, and the smallest is $\hat{b}_\mathrm{A} = 0.5565$ on Aug 8, 2011. The former is much bigger and the latter is much smaller than b_0. The number of iterations required also varies from case to case. It is e.g. just 3 on Aug 1, but 15 on Aug 9, 2011. The observed durations, the estimated diurnal patterns, $\hat{\phi}(t)$, and the standardized durations, $\hat{y}_i$, on Aug 2 and 8, 2011, respectively, are shown in Figure 1. The diurnal pattern on Aug 2, 2011 has a typical inverse U-form except for a sub-peak at about one hour prior to the close. In this case the smoothing quality of $\hat{\phi}(t)$ is very good. The diurnal pattern on Aug 8, 2011 looks very special with two peaks in the morning and in the afternoon and some minor structures. Hence $\hat{b}_{\mathrm{A}}$ is now very small. The bandwidths used around lunch time and prior to the close seem to be not large enough. This problem might be solved using a varying bandwidth taking $\hat{m}''(t)$ into account. Discussion on this is beyond the aim of this paper. The standardized durations seem to be stationary with clear clusters, except for a few outliers. Two EACD(1, 1) (exponential ACD) models are fitted to $\hat{y}_i$ in Figures 1(c) and 1(f), respectively, using the "fACD" package in R (Perlin, 2009). The results are listed in Table 2. These models fit the data very well. The estimation quality

Table 2. EACD parameters for the 1. (left) and 2. (right) examples with p-values

ω	α	β	ω	α	β
.0362(.0030)	.1197(.0053)	.8552(.0067)	.0487(.0028)	.1120(.0039)	.8473(.0054)

after removing $\phi(t)$ is improved. For instance, the EACD model obtained from x_i on Aug 2, 2011, is non-stationary with $\hat{\alpha} + \hat{\beta} \approx 1$. However, the first model given in Table 2 estimated from $\hat{y}_i$ does not share this problem.

5 Final remarks

A local linear estimator of diurnal duration patterns on financial markets with special asymptotics is proposed. An IPI bandwidth selector is developed. Applications show that the proposal works well in practice. To our knowledge this paper is the first effort to develop a data-driven estimator of diurnal duration patterns. There are still many open questions in this context, such as further improvements of the data-driven algorithm, the estimation of $m(t)$ via $\hat{f}(t)$ and the significance test of $\hat{\phi}(t)$. Furthermore, discussion on the application of the estimation results and detailed economic sources for a non-typical diurnal duration dynamics like those happened on Aug 8, 2011 is also of great interest.

Acknowledgment

We are grateful to Mr Christian Peitz and Mr Kan Wang for preparing the data, useful discussion and further application of the proposed algorithm.

References

Bauwens, L. and Giot, P. (2000). The logarithmic ACD model: An application to the bid/ask quote process of two NYSE stocks. *Ann. d'Econ. et de Stat.*, 60, 117-149.

Beran, J. and Y. Feng (2002). Iterative plug-in algorithms for SEMIFAR models - definition, convergence and asymptotic properties. *J. Comput. Graph. Stat.* 11, 690-713.

Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *J. Time Ser. Anal.*, 17, 247-270.

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

Engle, R.F. and J.R. Russell (2010). Analysis of high frequency data. In *Handbook of Financial Econometrics*, Vol. 1, 383-426, ed. by Y. Ait-Sahalia and L.P. Hansen, Elsevier.

Feng, Y. and Wang, K. (2011). Modelling long term and conditional dynamics in average durations by a semiparametric ACD. Forthcoming preprint, University of Paderborn.

Gasser, T., A. Kneip & W. Köhler (1991) A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.* 86, 643–652.

Hautsch, N. (2004). *Modelling irregularly spaced financial data.* Springer, Berlin.

Perlin, M. (2009). Autoregressive conditional duration models in R. On R-Forge.

Veredas, D., Rodriguez-Poo, J., and Espasa, A. (2002). On the (intradaily) seasonality, dynamics and durations zero of a financial point process. CORE Disc. Paper, 2002/23.

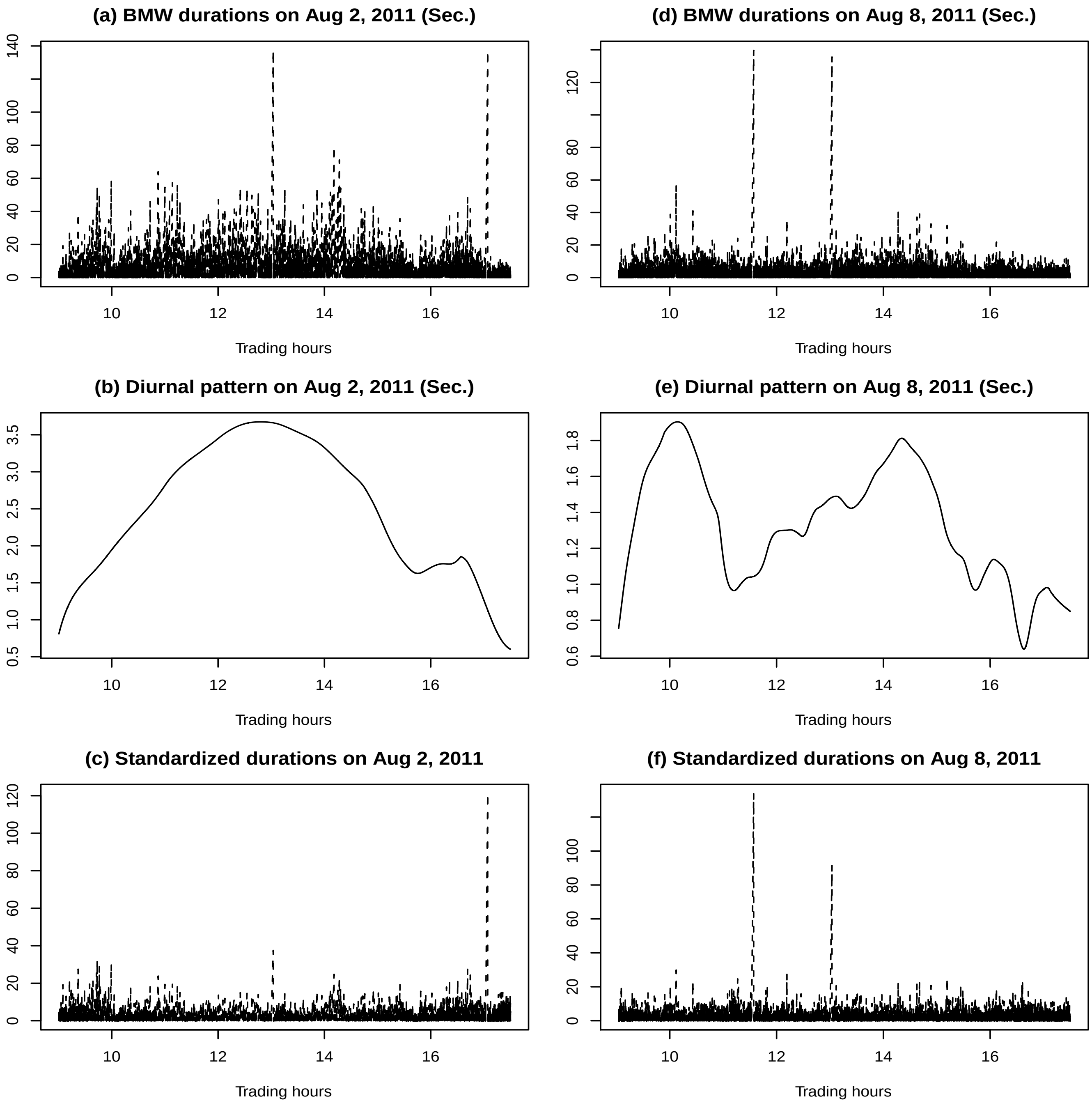

Figure 1: Observed durations (above), estimated diurnal patterns (middle) and standardized durations (bottom) of BMW stock on two days.

Recent discussion papers

2011-11	Yuanhua Feng	Data-driven estimation of diurnal duration patterns
2011-10	B. Michael Gilroy Birke Thuy Duong Nguyen	Ist Fairer Handel Wirklich Fair?
2011-09	Yuanhua Feng Zhichao Guo Christian Peitz Xiangyong Tan	A tree-form constant market share model for growth causes in international trade based on multi-level classification
2011-08	Tim Krieger Daniel Meierriecks	Ökonomische Aspekte des neuen globalen Terrorismus
2011-07	Tim Krieger Daniel Meierrieks	Terrorist Financing and Money Laundering
2011-06	Thomas Gries Stefan Jungblut Wim Naudé	New Firm Creation and Failure: A Matching Approach
2011-05	Daniel Kaimann	"To infinity and beyond!" - A genre-specific film analysis of movie success mechanisms
2011-04	Yuanhua Feng Zhichao Guo Christian Peitz Xiangyong Tan	A tree-form constant market share analysis for modelling growth causes in international trade
2011-03	Zhichao Guo Yuanhua Feng Xiangyong Tan	Impact of China's accession to WTO and the financial crisis on China's exports to Germany
2011-02	Alexander Haupt Tim Krieger Thomas Lange	Competition for the International Pool of Talent: Education Policy and Student Mobility
2011-01	B. Michael Gilroy Daniel Kruse	Die Prinzipal-Agent-Theorie als Erklärungsinstrumentarium von Korruption: Angewendet auf den Praxisfall „Siemens"
2010-09	Yuanhua Feng	An iterative plug-in algorithm for decomposing seasonal time series using the Berlin Method
2010-08	Zhichao Guo Yuanhua Feng Xiangyong Tan	Short- and long-term impact of remarkable economic events on the growth causes of China-Germany trade in agri-food products *[published in: Economic Modelling 28 (2011), 2359-2368]*
2010-07	B. Michael Gilroy Elmar Lukas Christian Heimann	Welchen Einfluss hat die Anwesenheit von ausländischen und multinationalen Unternehmungen auf die deutschen Exporte?
2010-06	Stefan Gravemeyer Thomas Gries	Income and disparity in Germany and China
2010-05	Thomas Gries Margarete Redlin	Short-run and Long-run Dynamics of Growth, Inequality and Poverty in the Developing World
2010-04	Stefan Gravemeyer Thomas Gries	Poverty in Shenzen

	Jinjun Xue	
2010-03	Alexander Haupt Tim Krieger Thomas Lange	A Note on Brain Gain and Brain Drain: Permanent Migration and Education Policy
2010-02	Sarah Brockhoff Tim Krieger Daniel Meierrieks	Ties That Do Not Bind (Directly): The Education Terrorism Nexus Revisited
2010-01	Claus-Jochen Haake, Tim Krieger, Steffen Minter	On the institutional design of burden sharing when financing external border enforcement in the EU
2009-06	Tim Krieger, Stefan Traub	Wie hat sich die intragenerationale Umverteilung in der staatlichen Säule des Rentensystems verändert? Ein internationaler Vergleich auf Basis von LIS-Daten *[published in: Jahrbücher für Nationalökonomie und Statistik 231 (2011), 2, 266-287]*
2009-05	Karin Mayr, Steffen Minter, Tim Krieger	Policies on illegal immigration in a federation *[forthcoming in: Regional Science and Urban Economics]*
2009-04	Tim Krieger, Daniel Meierrieks	Terrorism in the Worlds of Welfare Capitalism *[published in: Journal of Conflict Resolution 54, 6, 902-939.]*
2009-03	Alexander Haupt, Tim Krieger	The role of mobility in tax and subsidy competition
2009-02	Thomas Gries, Tim Krieger, Daniel Meierrieks	Causal Linkages Between Domestic Terrorism and Economic Growth *[forthcoming in: Defense and Peace Economics]*
2009-01	Andreas Freytag, Jens J. Krüger, Daniel Meierrieks, Friedrich Schneider	The Origin of Terrorism - Cross-Country Estimates on Socio-Economic Determinants of Terrorism
2008-11	Thomas Gries, Magarete Redlin	China's provincial disparities and the determinants of provincial inequality *[published in: Journal of Chinese economic and business studies 7 (2009), 2, 259-281.]*
2008-10	Thomas Gries, Manfred Kraft, Daniel Meierrieks	Financial Deepening, Trade Openness and Economic Growth in Latin America and the Caribbean *[forthcoming in: Applied Economics]*
2008-09	Stefan Gravemeyer, Thomas Gries, Jinjun Xue	Discrimination, Income Determination and Inequality – The case of Shenzen *[forthcoming in: Urban Studies]*
2008-08	Thomas Gries, Manfred Kraft, Daniel Meierrieks	Linkages between Financial Deepening, Trade Openess and Economic Development: Causality Evidence from Sub-Saharan Africa *[published in: World Development 37 (2009), 1849-1860.]*
2008-07	Tim Krieger, Sven Stöwhase	Diskretionäre rentenpolitische Maßnahmen und die Entwicklung des Rentenwerts in Deutschland von 2003-2008 *[published in: Zeitschrift für Wirtschaftspolitik 58 (2009), 1, 36-54.]*

2008-06	Tim Krieger, Stefan Traub	Back to Bismarck? Shifting Preferences for Intragenerational Redistribution in OECD Pension Systems
2008-05	Tim Krieger, Daniel Meierrieks	What causes terrorism? *[published in: Public Choice 147 (2010), 1-2, 3-27.]*
2008-04	Thomas Lange	Local public funding of higher education when students and skilled workers are mobile *[published in: Finanzarchiv 65 (2009), 2, 178-199.]*
2008-03	Natasha Bilkic, Thomas Gries, Margarethe Pilichowski	Stay at school or start working? - Optimal timing of leaving school under uncertainty and irreversibility
2008-02	Thomas Gries, Stefan Jungblut, Tim Krieger, Henning Meier	Statutory retirement age and lifelong learning
2008-01	Tim Krieger, Thomas Lange	Education policy and tax competition with imperfect student and labor mobility *[published in: International Tax and Public Finance 17 (2008), 6, 587-606.]*
2007-05	Wolfgang Eggert, Tim Krieger, Volker Meier	Education, unemployment and migration *[published in: Journal of Public Economics 94 (2010), 5-6, 354-362.]*
2007-04	Tim Krieger, Steffen Minter	Immigration amnesties in the southern EU member states - a challenge for the entire EU? *[published in: Romanian Journal of European Studies 5-6/2007, 15-32.]*
2007-03	Axel Dreher, Tim Krieger	Diesel price convergence and mineral oil taxation in Europe *[published in: Applied Economics 42 (2010), 15, 1955-1961]*
2007-02	Michael Gorski, Tim Krieger, Thomas Lange	Pensions, education and life expectancy
2007-01	Wolfgang Eggert, Max von Ehrlich, Robert Fenge, Günther König	Konvergenz- und Wachstumseffekte der europäischen Regionalpolitik in Deutschland *[published in: Perspektiven der Wirtschaftspolitik 8 (2007), 130-146.]*
2006-02	Tim Krieger	Public pensions and return migration *[published in: Public Choice 134 (2008), 3-4, 163-178.]*
2006-01	Jeremy S.S. Edwards, Wolfgang Eggert, Alfons J. Weichenrieder	The measurement of firm ownership and its effect on managerial pay *[published under the title “Corporate Governance and Pay for Performance: Evidence from Germany” in: Economics of Governance 10 (2009), 1, 1-26.]*